Harnessing Machine Intelligence in Automatic Echocardiogram Analysis: Current Status, Limitations, and Future Directions

Ghada Zamzmi[1], Li-Yueh Hsu[2], Wen Li[2], Vandana Sachdev[2], and Sameer Antani[1]

Abstract—Echocardiography (echo) is a critical tool in diagnosing various cardiovascular diseases. Despite its diagnostic and prognostic value, interpretation and analysis of echo images are still widely performed manually by echocardiographers. A plethora of algorithms has been proposed to analyze medical ultrasound data using signal processing and machine learning techniques. These algorithms provided opportunities for developing automated echo analysis and interpretation systems. The automated approach can significantly assist in decreasing the variability and burden associated with manual image measurements. In this paper, we review the state-of-the-art automatic methods for analyzing echocardiography data. Particularly, we comprehensively and systematically review existing methods of four major tasks: echo quality assessment, view classification, boundary segmentation, and disease diagnosis. Our review covers three echo imaging modes, which are B-mode, M-mode, and Doppler. We also discuss the challenges and limitations of current methods and outline the most pressing directions for future research. In summary, this review presents the current status of automatic echo analysis and discusses the challenges that need to be addressed to obtain robust systems suitable for efficient use in clinical settings or point-of-care testing.

Index Terms—Echocardiography, Ultrasound, Doppler, Cardiovascular Diseases, 2D Echo, Supervised Learning, Unsupervised Learning, Deep Learning, Image Processing, Echo Datasets, Point-of-care Testing

◆

1 INTRODUCTION

CARDIOVASCULAR disease (CVD) is the leading cause of mortality in the United States and globally [1]. CVD is diagnosed using several imaging techniques: echocardiography (echo), cardiac magnetic resonance imaging (CMR), multiple gated acquisition scan (MUGA), and computed tomography (CT). Of these techniques, echo is the most commonly used as it is non-invasive, portable, inexpensive, and widely available [2]. Transthoracic echocardiogram (TTE), a very safe and common type of echocardiogram, involves using a transducer to transmit ultrasound waves to the heart and converting the reflected waves (echoes) into images. The recorded echo data can be either a single shot (static image) at a specific cardiac period or a video sequence over cardiac cycles. A single cardiac cycle starts with ventricular contraction (systole) and ends by ventricular relaxation (diastole). Different echo modes can be obtained using TTE [2], namely M-mode, B-mode, and Doppler, each with purpose-specific characteristics. These modes are typically used in an integrated fashion to provide better visualization and diagnosis of various cardiac conditions. Descriptions of echo modes can be found in Appendix A.

Existing approaches for analyzing echo data can be broadly divided into manual and automated. In the manual approach, echocardiographers manually select good-quality end-systole and end-diastole frames followed by delineating the desired region and measuring cardiac indices. Examples of common cardiac indices include ejection fraction (B-mode), peak velocity (spectral Doppler), and posterior wall thickness (M-mode). Complete list of cardiac indices can be found in [2]. This manual approach has three limitations. First, it is error-prone and suffers from high intra- and inter-reader variability [3], [4]. Manual estimation of cardiac indices is more challenging and prone to larger variability in case of fetuses/infants [5] and animals [6] due to their small cardiac size and unclear boundaries. Second, the manual delineation is a tedious task requiring a significant amount of time. This time commitment paired with insufficient access to technicians increases the workload, which might lead to fatigue and distraction, and therefore, inaccurate or delayed diagnoses [7]. Third, cardiological expertise is a heavily burdened resource and often unavailable in low-resource settings.

Automated echo analysis systems can provide a timely, less subjective, and inexpensive alternative to the manual approach. Such systems can control intra- and inter-reader variability, greatly reduce the workload, and address the shortage of cardiological expertise in

---

[1] G. Zamzmi and S. Antani are with National Library of Medicine (NLM), National Institutes of Health (NIH), MD, USA.
[2] W. Li, L. Hsu, V. Sachdev are with National Heart, Lung, and Blood Institute, National Institutes of Health (NIH), MD, USA.

low-resource settings. This paper provides a comprehensive and systematic review of existing automated methods for four major echo tasks, namely quality assessment, mode/view classification, segmentation, and CVD diagnosis. The review covers three clinically used echo imaging modes, which are B-mode, Doppler, and M-mode. Previous reviews focus on other modalities (e.g., MRI), single mode (B-mode), specific task (e.g., segmentation), or algorithm (e.g., deep learning).

For example, Litjens et al. [8] presents existing deep learning algorithms for analyzing CT and echo modalities. The paper focuses mainly on convolutional neural network (CNN for classification) and fully convolutional neural network (FCN for segmentation) applied to B-mode images. Similarly, Meiburger et al. [9] reviews existing FCN segmentation methods applied to B-mode ultrasound images of the heart, abdomen, liver, gynecology, and prostate. Other reviews of automated segmentation methods applied to MRI and CT can be found in [10], [11], [12]. A more focused review of segmentation methods applied to B-mode fetal echocardiography is presented in [13]. For CVD diagnosis, Alsharqi et al. [14] presents machine learning methods applied to B-mode echo for disease classification. Similarly, Sudarshan et al. [15] presents a review of machine-learning method applied to 2D echocardiography (classification) and provide a summary of the most commonly used features for identifying a specific cardiac disease (infarcted Myocardium tissue characterization).

Contrary to previous reviews, this paper presents the first comprehensive and systematic review of automated methods for major echo tasks. It makes the following contributions:

- It presents the current status and challenges of existing automated echo analysis methods (Section 2).
- It provides a summary of the metrics used to assess the performance of various echo tasks (Section 3).
- It systematically and comprehensively reviews existing automated methods covering four major tasks: echo quality assessment (Section 4.2), view/mode classification (Section 4.3), segmentation (Section 4.4), and CVD diagnosis (Section 4.5).
- The review covers all echo modes, namely B-mode, M-mode, and Doppler. It also provides a summary of the most commonly used clinical and nonclinical features for identifying different CVD from different echo modes.
- It presents descriptions of existing publicly available echo datasets (Section 5).
- It highlight the most pressing directions for future research (Section 6).

Section 7 concludes the paper.

2 BACKGROUND

2.1 Echo Analysis: Artifacts and Challenges

The quality of echo data depends highly on the scanning technique and configurations. Because most of echo artifacts occur as a result of improper configurations and acquisition, echo images of a specific cardiac tissue acquired by different operators/vendors or under different configurations can have different visual appearances. These variations can confuse cardiologists and make the image interpretation task challenging. Examples of the main artifacts in B-mode and M-mode echo are: side lobe artifact, mirroring artifact, refraction artifact, and shadowing artifact [16]. The main artifacts of Doppler echo are: aliasing, mirroring, spectral broadening, and blooming [16]. A robust automated echo image analysis system should consider these variations and artifacts by including a variety of dataset collected from different machine configurations and operator setting for the algorithm development and model training.

Another major challenge of echo analysis is the presence of speckle noise. Speckle noise, which has a granular appearance, is a multiplicative noise that occurs when several waves of the same frequency and different phases and amplitude interfere with each other. This type of noise can greatly degrade the quality of the image, and therefore, the quality of the automated algorithms. Several despeckling techniques have been proposed to reduce the effect of speckle noise while preserving structure and contextual features as well as other useful information. We refer the reader to [17], [18], [19], [20] for reviews of traditional despeckling techniques and [21], [22] for deep learning-based despeckling techniques.

2.2 Current Status of Automated Echo Analysis

Existing automated echo analysis systems perform one of the following tasks:

- *Quality assessment* provides a quality score for an echo frame in real-time or classifies the acquired echo frame as low-quality (unmeasurable) or good-quality. Automating this task facilities the analysis of subsequent tasks because it automatically removes unmeasurable echo cases. For example, automated quality assessment can be used to exclude low-quality B-mode images with unclear boundaries or unmeasurable Doppler images with overlapped peaks (e.g., E and A peaks of mitral valve flow are overlapped).
- *Mode/View classification* is the categorization of acquired echo data into different modes (B-mode, M-mode, Doppler) or cardiac views. Each mode of echo can be recorded from different views. For example, a comprehensive B-mode acquisition involves imaging the heart from different windows or views by positioning the

transducer in different locations [2]. The most common B-mode views include [2]: Parasternal Long Axis and Short Axis views (PLAX and PSAX), Apical Two-chamber view (A2C), Apical Three-chamber view (A3C), Apical Four-chamber view (A4C), Apical Five-chamber view (A5C), Subcostal Long and Short Axis Views (SCLX and SCSX), and Suprasternal Notch View (SSN). Similarly, Doppler can be acquired, using continuous wave (CW) or pulsed wave (PW), from different locations to measure the function of different valves (e.g., aorta valve [AV], mitral valve [MV]). This task can greatly enhance subsequent tasks because it allows view-specific segmentation and diagnosis.

- *Boundary segmentation* task involves delineating the boundary or segmenting the area of a desired region. This region can be a cardiac chamber in B-mode images, wall in M-mode images, or a spectral envelope in Doppler images. The segmented region is then used to extract features or cardiac indices followed by CVD classification. Automating this task provides fast, accurate, and objective segmentation over the whole cardiac cycle with a minimum time cost.
- *CVD classification* is the detection or predication of specific cardiac disease based on image features or calculated cardiac indices. Fully automated machine-assisted or machine-based screening and diagnostic systems have a significant potential in providing high-quality and cost-efficient health care to the patients in low-resource settings.

The first step for all above-mentioned tasks is the detection or localization of the region of interest (ROI). Accurate ROI detection is an important step that increases the performance and decreases the computational complexity of the method because it removes irrelevant regions that confuses the algorithm. ROI detection in case of B-mode images involves cropping the anatomical area from the background (e.g., waveforms and texts) while the ROI detection in Doppler images involves cropping the Doppler signal region. As shown in the tables (Table 2 – Table 5), the majority of existing works manually localize ROI in echo images prior to further analysis. Other works use semi-automated or fully automated methods to detect ROI in echo images. However, these fully automated methods are view-specific and built with specific assumptions (e.g., distinct chamber shape or fixed locations of Doppler signal), and hence, might fail if these assumptions are violated.

3 EVALUATION METRICS

This section summarizes different metrics used to evaluate the performance of different echo tasks. We broadly divide these metrics into classification evaluation metrics and segmentation evaluation metrics.

3.1 Classification Evaluation Metrics

Classification metrics are derived from the confusion matrix, which shows the number of correct and incorrect classifications as compared to the ground truth labels [23]. Examples of derived metrics include accuracy, error rate (ER), true positive rate (TPR), and true negative rate (TNR).

Accuracy represents the number of instances (e.g., images or pixels) that are correctly classified divided by the total number of instances in the dataset ($\frac{TP+TN}{TP+TN+FP+FN}$), where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. ER measures the percentage of incorrect classifications. Dividing the number of instances that are incorrectly classified by the total number of instances gives ER; i.e., subtracting the accuracy percentage from 100. TPR (a.k.a., recall or sensitivity) measures the percentage of actual positive examples that are correctly classified. TNR (a.k.a., specificity) measures the percentage of actual negatives that are correctly classified. ROC (Receiver Operating Characteristic) curve [23] is another evaluation metric that is commonly used in medical applications. ROC plots the false positive rate (FPR) on X-axis and TPR on Y-axis at different threshold settings of the classifier. A curve that climbs toward the top-left corner indicates an ideal classification performance. The area under the ROC curve, known as AUC, is used to measure the quality of the classification models. The value of AUC ranges from 0 (worst) to 1 (best).

3.2 Segmentation Evaluation Metrics

Roughly, segmentation evaluation metrics can be classified as similarity-based metrics, distance-based metrics, and statistical-based metrics.

Similarity-based metrics measure the similarity between the automatically segmented region and the manually segmented region. This region can be left ventricle (LV) cavity in B-mode images or spectral envelope in Doppler images. Examples of the most common similarity-based metrics include Jaccard similarity index (JSI) and Dice similarity index (DSI). JSI evaluates the segmentation performance using TP, FP, and FN rates as follows [24]: $JSI = \frac{TP}{TP+FP+FN}$, where TP represents the pixels that are correctly classified as the target cardiac region, FP represents the background pixels that are falsely classified as the target region, and FN represents cardiac pixels that are falsely classified as background. DSI is another similarity-based metric that measures the similarity or intersection between the automatically labeled pixels and manually labeled pixels. Mathematically, this metric is formulated as follows [24]: $DSI = \frac{2 \times TP}{2TP+FP+FN}$. The main difference

TABLE 1
Strengths and limitations of algorithm categories

Algorithm Category	Strengths	Limitations
Low Level Image Processing (e.g., Thresholding and Edge Detection)	Simple implementation low computational complexity	Sensitive to the image's noise and artifacts Perform poorly when applied to obscured images and images with unclear boundaries, non-uniform regional intensities, and confusing structures
Deformable Models (e.g., Active Contour Model)	Can segment any shape Highly flexible	Sensitive to the initial contour location/shape Perform poorly when the shape vary widely Tend to become computationally complex
Statistical Models (e.g., Active Appearance Model)	Use intensity and shape information Highly effective	Require proper initialization Expensive manual shapes annotations Perform poorly when the shape vary widely Local minimum trap
Conventional Machine Learning (e.g., Random Forest Trees)	Good to high performance Good interpretability	Look at specific handcrafted features Bias of engineer who designs the method Require a set of annotated data
Deep Learning (e.g., Convolutional Neural Network)	Superior performance	Require a large set of annotated data Long tuning/training process Lack of interpretability

between JSI and DSI is that DSI counts TP twice while JSI counts TP once. The value of both JSI and DSI ranges from 0 to 1, where 0 indicates complete dis-similarity and 1 indicates complete similarity. Intersection over Union (IoU) is another metric that calculates the intersection between two regions by dividing the area of overlap between them by the area of union.

A single evaluation category can perhaps not be enough to evaluate the performance of a segmentation algorithm. In addition, similarity-based metrics only report the degree of overlapping and do not report or consider the location or distance between the segmentation and ground truth. Distance-based metrics, on the other hand, consider how far apart the segmentation and ground truth are from each other. Average Contour Distance (ACD) and Average Surface Distance (ASD) are two distance-based metrics that are commonly used to evaluate regions segmentation. Both ACD and ASD are measured in millimeter (mm) [24].

Statistical-based metrics are used to measure the correlation between the automatic and manual segmentation. Specifically, statistical-based metrics evaluate the accuracy of segmentation by measuring the correlation between the cardiac indices calculated based on the automatic segmentation and the manual indices. Correlation coefficients (CC) and Bland-Altman agreement (B&A) are two important statistical metrics that are commonly used to evaluate the performance of cardiac segmentation. CC measures the correlation or the agreement between two sets of data. The mathematical formula of CC, which can be found in [25], returns a value that ranges from -1 to 1, where 1 indicates a strong positive correlation, -1 indicates a strong negative correlation, and 0 indicates no correlation. B&A measures the agreement between two set of measurements or data using the mean difference and limits of agreement. The mathematical formulation of B&A and comparison with CC metric can be found in [26].

4 AUTOMATED ECHO ANALYSIS

As computing technology and machine intelligence algorithms evolve, automated analysis of echocardiograms have the potential to improve clinical workflows and enhance diagnostic accuracy. This section provides a comprehensive review for existing automated methods of four tasks: *echo quality assessment*, *mode/view classification*, *boundary segmentation*, and *CVD classification*. The automated methods of these tasks can be divided, based on the underlying algorithm, into low level image processing-based methods, deformable model-based methods, statistical model-based methods, conventional machine learning-based methods, and deep learning-based methods. Table 1 summarizes the advantages and disadvantages of these five algorithm categories.

4.1 Literature Review Design

To ensure the reproducibility of this review, we present our search and selection strategies. A flowchart of our literature review is depicted in Figure 1.

4.1.1 Search Strategy

We did a systematic review of automated echocardiography using PubMed, IEEE Xplore, Google Scholar, Google Datasets, ACM Digital Library, CiteSeer, PLOS ONE, and Scopus search engines. We searched for scientific conferences, journal articles, technical reports, and dataset papers published up to February 2020, and retrieved relevant literature by using a combination of keyword terms. Examples of these terms include cardiac imaging; automated echo analysis and interpretation; echo review/survey; echo mode/view classification; echo disease classification; echo quality assessment; image-based analysis echo; machine learning-based analysis echo; deep learning-based analysis echo; chamber/envelope/wall segmentation echo; and echocardiographic datasets. Terms related to echocardiographic
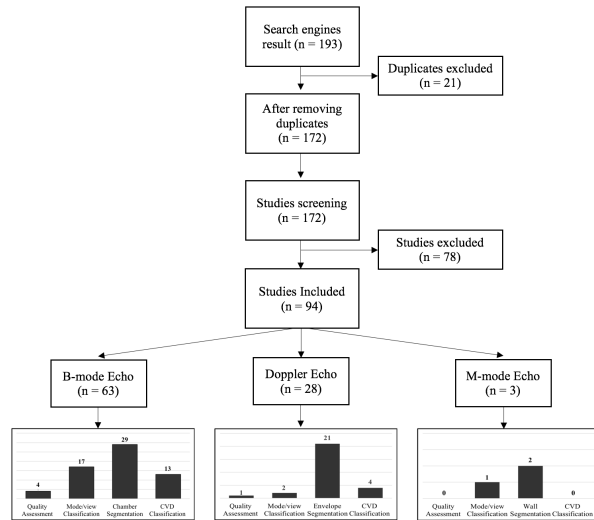
Fig. 1. Flowchart of our review. The histogram associated with each mode represents the number of automated works for each task.

hardware and other cardiac imaging modalities (e.g., CT and MRI) are excluded because they are outside the scope of this review. We retrieved, using this search strategy, a total of 193 studies.

4.1.2 Selection Strategy

We included a study if all of the following criteria are fulfilled: (1) the full text is written in English; (2) the study includes a clear description of the technical method and used dataset; (3) the study is published as a full conference paper, journal article, open access article, or technical report; and (4) the study is published the year of 2004 or after because a rising amount of interest and publications in automated echocardiography analysis using image processing and machine learning sprouted around that time. We screened the retrieved papers independently and excluded the ones that failed to adhere to these criteria. We included, using this strategy, a total of 94 papers in this systematic review. The selected papers are loaded into EndNote X8 and categorized into different groups.

4.2 Quality Assessment

Unlike other cardiac imaging modalities, the diagnostic accuracy of echocardiography is highly dependent on the image quality at the acquisition stage. Therefore, the quality of the acquired echo depends highly on the technician's expertise. Automated echo quality assessment task provides a quality score of a given image or categorizes this image as low- or good-quality. These methods can aid during echo acquisition by providing real-time feedback and automatically rejecting low-quality cases. We divide existing automated quality assessment methods into two categories: model-based methods and deep learning-based methods. Table 2

summaries current automated methods for echo quality assessment.

4.2.1 Model-based Methods

One of the first automated methods for assessing echo quality is presented in [27]. The proposed method models the four chambers (left ventricle [LV], right ventricle [RV], left atrium [LA], right atrium [RA]) of A4C view by a non-uniform rational B-splines (nUrBs) using 12 control points. Then, the nUrBs models for all chambers are joined by similarity transforms to create a complete view model. Finally, the model goodness-of-fit is used to calculate a quality score. The proposed method is tuned using 35 B-mode (A4C) echo videos recorded from 4 healthy volunteers. The recorded videos include both good quality and completely erroneous quality. Each of the recorded video is scored as having good, fair, or poor quality by 2 cardiologists. The proposed method improved the quality of the recorded A4C images from poor to fair or good by 89% (i.e., 8 of 9 cases were improved).

Another B-mode echo quality assessment method is presented in [28]. The presented method assesses the quality by comparing the structure of a representative atlas (model) with the structure of the input image. The structure of PLAX atlas is generated from 89 manually segmented images while the structure of the input image is generated using thresholding and the Generalized Hough Transform (GHT). The proposed method is evaluated using echo data (133 PLAX images) of 35 normal and hypertrophic patients. Each image is scored by an expert sonographer as poor, moderate, and good visibility. The automatically generated scores achieved good correlation with manual ratings (correlation coefficient = 0.84).

Although model-based methods for echo quality assessment can achieve good performance, these methods are view-specific because they require to generate a specific model or template for each view. In addition, the accurate generation of the template relies heavily on human experts or the image's contrast. For example, methods of Snare et al. [27] and Pavani et al. [28] are designed for a specific B-mode view (A4C [27] or PLAX [28]), require manual annotation [28], and both rely heavily on the presence of the sharp edges in the image; i.e., they would fail when applied to low contrast images.

4.2.2 Deep Learning-Based Methods

Abadi et al. [29] proposed a regression CNN architecture for assessing the quality of B-mode videos (A4C view). The proposed architecture is composed of two convolutional layers, each followed by Rectified Linear Units (ReLU), two pooling layers, and two fully connected layers. The loss function (L_2 norm) outputs the Euclidean distance of the network score to the manual

quality score. The proposed regression CNN architecture is trained using stochastic gradient descent (SGD), a batch size of 16, a momentum of 0.95, weight decay of 0.02, and initial learning rate of 0.0002. The architecture is trained using 2,344 end-systolic A4C frames. Evaluating the performance on 560 test set achieved a mean absolute error (MAE) of 0.87 ± 0.72.

Abadi et al. [30] extends their previous work [29] to include other cardiac views, namely A2C, A3C, A4C, PSAX at the aortic valve, and PSAX at the papillary muscle, as well as echo cine loops instead of static frames. The proposed multi-stream network architecture consists of five regression models with the same weights across the first few layers. The last layers of the proposed architecture are view-specific layers. Similar to [30], the loss function (L_2 norm) for each view computes the Euclidean distance of the network score to the manual quality score. The proposed architecture is trained using Adam optimizer and random initialization. This method, which is trained using 4,675 cine loops, achieved a mean quality score accuracy of $85\% \pm 12$ when applied to testing cine loops (1144).

In summary, there has been a little effort [27], [28], [29], [30] to create automated methods for B-mode echo quality assessment. In the case of M-model, we are not aware of any automated method for assessing the quality of the acquired M-mode images. As for Doppler, we are only aware of a recent deep learning-based method presented by Zamzmi et al. in [31]. The proposed method, which was trained on labeled images (good- and bad-quality) representing a wide range of real-world clinical variation, achieved 88.9% overall accuracy. We refer the reader to [31] for a detailed description of the method and presentation of the results.

Existing methods for assessing B-mode echo quality can be divided into model-based methods and deep learning-based methods. As shown in Table 2, deep learning-based methods [29], [30] achieved better performance as compared to model-based methods [27], [28]. The higher performance in deep learning-based methods could be attributed to a broader dataset exploited in the study [29], [30] as well as a more complex feature extraction and model learning. In addition, the deep learning-based method proposed in [30] is evaluated in a dataset collected from different US machines under different configurations in opposition to the methods presented in [27], [28]. Such setting for data collection ensures that the proposed method would be clinically relevant. Another advantage of deep learning-based methods is that these methods do not require the user to build a model or template for each view.

In the future, we would expect to see increasingly more deep learning methods to extend existing B-mode quality assessment methods, and to include quality assessment for all echo modes and views collected using multiple vendors under different configurations. Also, we would expect to integrate quality assessment task into acquisition software to provide a quality score for recorded echo frames in real-time.

4.3 View Classification

Mode or view classification is the categorization of echo images into different cardiac modes (e.g., B-mode) or views (e.g., A4C). Automating this task offers two main benefits. First, it facilities the organization, storage, and retrieval of echo images. Second, it is important for automating subsequent tasks. For example, measuring the function of a specific valve requires knowing the view beforehand because different views show different valves. We broadly categorize existing methods for mode/view classification into: conventional machine learning-based methods and deep learning-based methods. Table 3 provides a summary and quantitative comparisons of these methods.

4.3.1 Conventional Machine Learning-Based Methods

These methods use handcrafted features extracted from a detected ROI region with conventional machine learning classifiers to perform view classification. For example, Wu et al. [32] proposed a global approach that uses GIST descriptor with support vector machines (SVM) for classifying 8 B-mode views: PSAX, PLAX, A2C, A4C, SC4C, SC2C, SCLX, and other. GIST descriptor computes the spectral energy of the image and outputs a single feature vector. It uses blocks (4 pixels $\times$ 4 pixels) that contains several oriented Gabor filters to model the structure of the image. The final feature vector that represents the entire image is generated by moving these blocks over the image to generate spectrograms followed by concatenating the generated spectrograms. The extracted feature vectors for all images are used to train a probabilistic SVM. The proposed method achieved 98.51% overall accuracy when evaluated on a testing set. Other methods that use descriptors similar to GIST with SVM can be found in [33] (Scale-invariant feature transform [SIFT] descriptor) and [34] (histogram of oriented gradients [HOG] descriptor).

An earlier machine learning-based method for view classification is presented in [35]. The first stage of this method involves training LV detectors for four B-mode views (A4C, A2C, PLAX, PSAX) using a previous approach that incorporates Haar-wavelet type local features and boosting learning technique. Then, global templates (A4C/A2C template, PLAX template, and PSAX template) are constructed based on the detected LV regions and sent to multi-class classifiers. Each classifier is trained using the training images provided by its detector. The final classification is obtained by fusing the classes of all views. The proposed method achieved a classification accuracy over 96% when evaluated on a testing set. This method requires a consistent presence

of LV in all views, which limits its usage to cases that hold this constraint.

Instead of building individual LV detectors for each view, Ebadollahi et al. [36] proposed a method that detects the location of chambers using a generic detection approach (GSAT detector). The method models the spatial relationships among cardiac chambers to detect different views. For each view, the chambers spatial relationships and the statistical variations of their properties are modeled using Markov Random Field (MRF) relational graph. The method depends on the assumption that if any two images contain the same chambers where each chamber is surrounded by similar chambers, then the probability that these two images belong to the same view is high. Each model or "cardiac constellation" is assigned a vector of energies according to the different view-models. The energy vectors obtained from all the training images are used to build a multi-class SVM. Evaluating the proposed method using leave-one-video-out cross validation (LOOCV) achieved up to 88.35% average precision. The dataset that is used for training and testing the method contains 15 normal echo videos, 6 abnormal echo videos, and 10 B-mode views: 2 PLAX views, 4 PSAX views, and 4 apical views. The normal cases are used for training and testing, and the abnormal cases are used only for testing. A main limitation of this method includes sensitivity to ROI detector, noise, and image transformation.

The methods presented in [32], [33], [34], [35], [36] use spatial features extracted from static images instead of videos. In [37], Kumar et al. incorporates temporal or motion information with spatial features to classify 8 B-mode views: A2C, A3C, A4C, A5C, PLAX, PSAX, PSAP, and PSAM. The method starts by manually locating the ROI region in all videos followed by aligning (affine transform) these videos using the extreme corner points of the fan sector. Then, optical flow is applied to each frame to obtain the motion magnitude. Because motions in echo video are only useful when it is associated with the anatomical structures, the motion magnitude images are filtered using an edge map on image intensity. After obtaining the edge-filtered motion maps, several landmark points are detected using SIFT descriptor. Once the salient features are detected and encoded for each frame, the salient features of all frames in the training dataset are used to construct a hierarchical dictionary. This dictionary is used to train a kernel-based SVM. To classify a new input video, the trained classifier provides a label for each frame in the given video and used majority voting to decide the final class label of the video. The proposed method, which is trained using 113 videos, achieved 51%-100% correct classification rates when evaluated on a testing set. The main strength of this method is that it does not require constructing spatial and temporal models for each cardiac view. Other conventional-based methods for view classification are presented in [38] (bag of visual words with SVM),

[39] (gradient features and logistic trees), [40] (visual features and boosting), [41] (B-spline and thresholding), and [42] (histogram features and neural network).

Instead of using handcrafted features with traditional classifiers, convolutional neural network (CNN) can provide objective features extracted directly from the image at multiple level of abstractions while preserving the spatial relationship between the image pixels. These networks achieved state-of-the-art performance in different medical domains, including echocardiography.

4.3.2 Deep Learning-Based Methods

Recent works utilize CNN architectures for feature extraction and classification. Examples of CNN architectures that have been used for cardiac view classification include VGG [43], DenseNet [44], and ResNet [45].

For example, Zhang et al. [46] used VGG CNN [43] for distinguishing 6 B-mode views: A2C, A3C, A4C, PLAX, PSAX, and other. Prior to feature extraction and classification, each frame is converted into grayscale and re-sized (224×224). The re-sized image is then sent to VGG [43]. The output of the network is the view that has the highest probability of Softmax function. The entire network is trained using 40,000 pre-processed images with ADAM optimizer, 1×10^{-5} learning rate, mini-batch size of 64, and 20 epochs. Testing the trained network using cross-validation protocol achieved excellent accuracy (e.g., 99% accuracy for A4C views). Zhang et al. expanded their work in [47] to distinguish 23 different echo views. The codes and model weights for both works are available online [46], [47]. Similarly, Madani et al. [48] used VGG-based [43] method to distinguish 15 different echo views: 12 views from B-mode (e.g., PLAX and A4C), M-mode, and two Doppler views (CWD and PWD). The final layer of VGG-16 performs classification using Softmax function with 15 nodes. The network is trained using RMSprop optimization over 45 epochs. The overall test accuracy of distinguishing 12 views of B-mode images is above 97%. The accuracies of distinguishing CWD, PWD, and M-mode views are 98%, 83%, and 99%, respectively.

Another recent architecture for cardiac view classification is presented in [49]. The proposed cardiac view classification (CVC) architecture is designed and trained to distinguish 7 B-mode views, namely A2C, A4C, PLAX, PSAX, SC4C, SCVC, and ALAX, as well as unknown view. The pre-processing stage involves normalizing the image and down-sampling it to 128×128. The input image is then sent to a series of deep learning blocks where each block consists of convolution layer, max-pooling layer, inception layer, and concatenation layer. This network is trained over a maximum of 100 epochs using Adam optimizer and a mini-batch size of 64. Cross-entropy and mean absolute error (MAE) loss functions are used for computing the error and updating the weights, which are initialized using Uni-

form method. Ten fold patient-based cross-validation technique is used for training and validation (265,649 frames) while an independent set of unseen data (229,951 frames) is used for testing. The proposed CVC network achieved 97.4% and 98.5% overall accuracies for frame-level and sequence-level view classification.

A lightweight cardiac view classifier is introduced recently by Vaseli et al. [50] to distinguish 12 B-mode views. Several deep learning lightweight models are built based on three CNN architectures, VGG-16 [43], DenseNet [44], and ResNet [45]. These lightweight models contain approximately 1% of regular deep models' parameters and they are 6 times faster at run-time. The training parameters for all the deep and lightweight models can be found in [50]. These models are trained and evaluated using 16,612 echo videos collected from 3,151 patients. Combining the three lightweight models achieved an accuracy of 88.1% in classifying 12 cardiac views.

Instead of only extracting deep features from static images, Gao et al. [51] fused the output of spatial and temporal networks to classify the cardiac view of echo videos. The spatial CNN takes a 227×227 frame as input and sends the input frame to 7 convolutional layers for deep features extraction. The temporal CNN takes as input the acceleration image, generated by applying optical flow twice, and sends this image to 7 convolutional layers for feature extraction. Then, the output of spatial and temporal CNNs are fused together. The final view classification is obtained by the linear combination of both CNNs scores using Softmax function, which provides the probability of 8 classes (A2C, A3C, A4C, A5C, PLA, PSAA, and PSAP). Both networks are trained using a random initialization, 0.01 learning rate, and 120 epochs. Evaluating the proposed method on 152 echo videos achieved 92.1% accuracy. The accuracy of view classification using only the spatial CNN is 89.5%.

Table 3 provides a summary of automated view classification methods. As shown in the table, deep learning-based methods for view classification achieved excellent performance comparable to the human inter-observer performance, and outperform conventional methods in various views (e.g., A3C, A5C, PSAM, [37] vs [51]). In addition, deep learning-based methods are evaluated using larger datasets collected by different machines as compared to the conventional machine learning-based methods. These results suggest the superiority of deep learning-based methods in the presence of relatively large datasets. This indicates better generalizability of these methods across machines and settings. Deep learning methods, however, suffer from interpretability and transparency issues (black box).

To summarize, the majority of current automated methods focus on detecting different views of B-mode. Only a few works [31], [48] includes other echo mode such as M-mode and Doppler. Because automated view

classification is critical to obtain a fully automated and real-time system that can be used efficiently in clinical practice, there is a need for future research focus on developing automated and lightweight view classification for all echo modes (B-mode, M-mode, and Doppler).

4.4 Boundary Segmentation

The current practice for cardiac boundary segmentation requires technicians to perform manual delineation followed by using the traced boundaries for computing structural and functional indices. This practice is tedious, error-prone, and subject to high intra- and inter-readers variation. In this section, we review automated methods for segmentation in B-mode, Doppler, and M-mode images, and provide a summary in Table 4.

4.4.1 B-mode, Chamber Segmentation

We categorized the methods of chamber segmentation into five categories (Table 1): low level image processing-based methods, deformable model-based methods, statistical model-based methods, conventional machine learning-based methods, and deep learning-based methods.

Low Level Image Processing-based Methods: Melo et al. [52] proposed a low level image processing-based method for segmenting LV chamber. The proposed method has two main modules: pre-processing module and segmentation module. The pre-processing module takes a raw image, performs filtering and morphological operations, and sends the processed image to the segmentation module. This module uses watershed algorithm for segmenting LV border (PLAX view). After detecting LV border, several structural indices, such as LV area, are computed. The proposed method is evaluated using videos of 12 healthy volunteers and measured using eight different metrics [52]. Amorim et al. [53] uses a method similar to [52] for segmenting LV border in PLAX image. The main difference between [52] and [53] is that Amorim et al. [53] applies the watershed algorithm to a composite image obtained by combining the images of three cardiac cycles; this allows to exploit the similarity of corresponding frames from different cycles. As visually reported in [53], using the composite image led to increased delineation accuracy.

Instead of segmenting a specific chamber, John and Jayanthi [54] presented a low level image processing-based method for segmenting all cardiac chambers. The method starts by converting a 2D echo video (2 to 3 seconds) to grayscale frames. It then applies Speckle Reducing Anisotropic Diffusion (SRAD) filter to remove speckle noise from the image. To approximate the chamber locations, k-means algorithm is applied to create clusters of pixels with similar intensities followed by thresholding using an empirically determined value. Visual results demonstrated good agreement between the contour obtained by the proposed method and the

manual contour. This method fails to segment frames that have low contrast or dropouts on LV internal walls.

Other low level image processing-based methods for chamber segmentation can be found in [55] (watershed algorithm), [56] (Otsu thresholding and edge detection), [57] (thresholding and morphological operations), [58] (watershed algorithms), and [59] (thresholding and morphological operations). The methods of this category are easy to implement and have low computational complexity as compared to the methods of other categories. However, these methods are highly sensitive to the signal-to-noise ratio (SNR). In addition, these methods perform poorly, and might completely fail, in detecting the border in images with obscure boundaries, non-uniform regional intensities, and confusing anatomical structures (e.g., valve).

Deformable Model-Based Methods: Chen et al. [60] generates the active contour of LV by solving a coupled optimization function that combines shape and intensity priors. The first optimization part is the weighted sum of the energy of the geometric contours of similar shapes. Minimizing this energy provides the initial contour and the transformation that aligns it to the prior shape. The geometric contours of all shapes, which are used to generate the prior shape, are obtained by manually tracing the cardiac boundaries of 85 images captured from 61 patients at end-diastole (ED). The second optimization part provides the optimal estimate of the weight by maximizing the mutual information of the image geometry (MIIG). The process of solving both parts generates the final LV segmentation. The visual results demonstrate that the proposed method can provide LV contours that are close to the contours provided by experts. It also shows that using MIIG provides a better description than MI (mutual information) because MIIG takes into account the neighborhood intensity distribution. MIIG, however, has a significant computational cost. A simpler active contour-based method is presented in [61]. The method combines Hough transform and active contour to detect LV in PSAX and PLAX images. Hough transform is used to generate LV initial shape. Active contour is then used to generate, via energy minimization, the final exact shape of LV. The detected LV border is used to calculate the following indices: LV areas in PSAX and PLAX views, LV volume, LV mass, and wall thickness.

Conventional active contour methods suffer from slow convergence. In [62], Marsousi et al. used B-spline snake algorithm for segmenting the endocardial boundary of LV chamber. The presented method does not require expensive optimization computation and is faster than conventional active contour methods. The main limitation of this method lies in the selection of the initial contour; i.e., if the selected initial contour lies far from the actual boundary, higher iterations of balloon force or Gradient Vector Flow [63] should be executed, which causes error and leads to tremendous increase in the time complexity. To avoid this problem, the method requires experts to manually select some points inside LV chamber. To automate the point selection, Marsousi et al. extends their method in [64] to select the best initial contour using a novel active ellipse model. Particularly, the intersection point of all chambers in A4C view is detected at the nearest point to the mass center. After detection the point, an initial ellipse is placed on the top-left side of the point followed by growing the initial ellipse until it fits the boundary. This method is tested using 20 A2C and A4C images collected from normal and abnormal cases. A comparison between this approach [64] and the previous approach [62] is performed using Dice's Coefficient (90.66 ± 5.17 [62] and 92.30 ± 4.45 [64]) and computational time (1.52 ± 0.82 [62] and 0.63 ± 0.29 [64]).

Other deformable-based methods for chamber segmentation can be found in [65] (Speckle resistant Gradient Vector Flow and B-spline), [66] (variational level set approach), [67] (k-means and active contour), [68] (constrained level-set), [69] (phase-based level set evolution), [70] (phase-based level set evolution), and [71] (active contour model and SIFT). Although deformable-based methods provide accurate segmentation, these methods are view-specific and hence do not perform well with widely varying shapes. In addition, these methods are highly sensitive to the initial contours and tend to become computationally complex.

Statistical Model-Based Methods: One of the first works that use statistical model for LV segmentation is presented in [72]. The proposed framework consists of three main stages: global despeckling, Active Appearance model (AAM) training, and LV segmentation. The global despeckling reduces speckle noise while maintaining the important image features. The second stage involves generating AAM model that represents the shape and texture of all end-diastole (ED) and end-systole (ES) images in the training set. To model the shape from the training images, the manually labeled contour and four landmark points in each image are used to align and register the images in the training set. The appearance model of AAM is generated using a weighted concatenation of three parts: the intensities of the original image shape, the intensities of the denoised image shape, and the mean gradient at each of the four landmark points. The final AAM model is constructed from the eigenvectors of the largest eigenvalues that are obtained by applying PCA to the combined model (shape and texture model). The third stage involves positioning the model in a new target image by solving an optimization problem. The proposed approach is tested using two fetal datasets: synthetic fetal echo images and clinical fetal echo images. The overall segmentation accuracies of the proposed method are 84.12% and 84.39% for synthetic and clinical images, respectively. The vi-

sual results demonstrate the superiority of the proposed method as compared to methods that use active shape models (ASM) [73], [74] as well as conventional AAM and constrained AAM [75].

Statistical based methods for chamber segmentation are view-specific and sensitive to the large variations in shape or appearance; i.e., cannot handle the large variations in chamber shape and appearance. Also, these methods can easily be trapped in local minima and require manual annotation.

Conventional Classification-Based Methods: The methods in this category utilize traditional machine learning approaches for labeling each pixel as chamber or background.

A machine learning-based method for fetal chambers segmentation is presented in [76]. The method starts by initializing a dictionary D_0 as a random matrix and computing the sparse coefficients of this matrix (X_0) from the training samples using Orthogonal Matching Pursuit (OMP). To generate a compact dictionary, sub-dictionaries (atoms) with utilization ratios less than a pre-determined threshold is discarded followed by updating the atom indices and coefficients to obtain a new group dictionary. After learning the group dictionary D, a new test sample is converted to two sparse coefficients X_{out} and X_{in} with respect to D_{out} and D_{in} sub-dictionaries, where out and in subscripts indicate the area outside and inside the chambers. The corresponding reconstruction residue R_{out} and R_{in} are then calculated using a proposed reconstruction residue function.

The final boundary is obtained by classifying each pixel in the sample image as one or zero using the calculated minimum reconstruction residue. The proposed method, called Adaptive Group Dictionary Learning, is evaluated using 40 clinical fetal echocardiograms. The experimental results demonstrate the efficiency of the proposed method as compared to previous machine learning-based methods [77], [78]. The construction of only two sub-dictionaries limits the proposed method to images that have two intensity patterns, and suggest that it might fail when applied to images with several intensity patterns.

Deep Learning-Based Methods: Semantic CNNs divide the image into different objects by labeling each pixel with the class of its enclosing object. These networks consist of only convolution and pooling layers organized in an encoder-decoder structure.

In [46], four separate semantic U-net models [79] are trained for segmenting the cardiac structures in PLAX, PSAX, A2C, and A4C views. The number of training data (images and masks) for each model are $128, 72, 168$, and 198 for PLAX, PSAX, A4C, and A2C, respectively. The training data for all models are augmented using cropping and blacking out techniques. All the models are trained using ADAM optimizer, 1×10^{-4} learning rate, 1×10^{-6} weight decay, 0.8 middle layers drop out, mini-batch size of 5, and 150 epochs. The trained models achieved good to excellent performance with IoU values that range from 73 to 92. The segmented cardiac chambers for each image is used to compute geometric dimensions, volumes, mass, longitudinal strain, and ejection fraction. These indices are then used for assessing cardiac structure and function. As discussed in the paper, the proposed automated framework showed superior performance as compared to manual measurements across all cardiac indices. Recent studies that use U-net and FCN for chamber segmentation can be found in [80], [81].

Machine learning-based methods (conventional and modern) for chamber segmentation showed excellent performance and outperformed the performance of human experts. However, building robust machine learning-based methods require a relatively large and well-annotated datasets. Also, these methods, especially deep learning-based, can be computationally expensive. Further, these methods may segment pixels outside the desired cardiac region due to the lack of model constraint. Finally, existing deep learning-based methods lack interpretability; i.e., they do not interpret nonlinear features or show the important human-recognizable clinical features.

4.4.2 Doppler, Envelopes Segmentation

The accurate tracing of spectral envelopes and estimation of maximum velocities in Doppler images has a great clinical significance. We review next existing methods for spectral envelope segmentation, and provide a summary in Table 4.

Low Level Image Processing-Based Methods: Zolgharni et al. [82] presented a thresholding-based method to detect spectral envelopes in long Doppler strips that span over several heartbeats. The analysis of long Doppler strips allows to extract additional velocity measures and leads to better understanding of the cardiac function. The method starts by manually locating the Doppler region (ROI) followed by converting pixel to velocity on the vertical axis and pixel to time on the horizontal axis. The baseline (zero velocity) is then determined and used to separate the negative Doppler profiles. Positive Doppler profiles are detected using a proposed objective thresholding method. The generated binary images are further processed to remove small connected areas. Finally, maximum velocity profiles are obtained using the biggest-gap algorithm as follows. A column vector is scanned from left to right to find a gap (cluster of consecutive black pixels). The largest gap from the top is selected as a point on the profile. The final output of the Biggest-Gap algorithm represents the maximum velocity envelope. This envelope is further smoothed using a low-pass first-order Butterworth filter. To extract Doppler indices from the spectral envelopes, Gaussian

model is fitted to the velocity profile and used to calculate peak velocity and velocity time integral. The automated measurements of velocity-time-integral showed strong correlation (r = 0.94) and good Bland-Altman agreement (SD = 6.9%) with the expert values. Similarly, the automated measurement of peak-velocities showed strong correlation (r = 0.98) with the expert values.

Another low level image processing-based method is presented in [83]. The proposed method extracts three Doppler indices, namely peak pressure gradient, peak velocity, and pressure half time, from the spectral envelopes for the purpose of assessing the severity of aortic regurgitation (AR). The method starts by locating the Doppler ROI based on the fixed locations of the vertical and horizontal axes (specific assumption). Before applying edge detector, several pre-processing operations such as noise filtering and contrast adjustment are performed. Then, Canny edge detector is applied to segment the spectral envelope. Once the envelope is segmented, the horizontal and vertical axes are converted into time and velocity. Finally, the curve is scanned to detect the highest peak value, which is used to compute the peak pressure gradient and pressure half time. To evaluate the performance of the proposed method, the automatic indices, computed from Doppler images of 11 subjects, are compared with human assessment. The results proved the feasibility of using the proposed algorithm in assessing the severity of AR as it showed strong correlation with human assessment for three age groups: 0.98 correlation for group 1 (20-35 years old), 0.92 correlation for group 2 (36-50 years old), and 0.83 correlation for group 3 (51-60 years old).

Texture analysis is a low level image operation that involves detecting regions in a given image based on their texture content (i.e., spatial variation in pixel intensities). Applying texture filters to an image returns a filtered image. Each pixel of this new image is a statistical representation of a neighborhood around this pixel in the original image. Biradar et al. [84] proposed to use combinations of three texture filters, which are entropy, range, and standard deviation, to detect the envelope in CW Doppler images. The filtered image is then thresholded and processed morphologically using erosion and dilation operations. The proposed method is evaluated using CW images of 25 patients suffering from aortic regurgitation. The experimental results showed that using a combination of entropy, range, and standard deviation filters can accurately delineate the spectral boundaries of CW Doppler images.

Other low level image processing-based methods for detecting velocity profiles can be found in [85] (thresholding and edge detection), [86] (Otsu thresholding), [87] (empirical thresholding and Random sample consensus), [88] (thresholding and edge detection), [89] (local adaptive thresholding), and [90] (thresholding and edge detection). Low level image processing-based methods can detect spectral envelopes of different blood flows without any pre-training and with a minimum amount of data. However, these methods are very sensitive to the image noise and artifacts as well as the image contrast and intensity patterns.

Deformable-Based Methods: Gaillard et al. [63] investigated the use of active contour for detecting Doppler spectral envelopes. The initial snake is generated using an automated method presented in [65]. The final snake of the envelope is then found by minimizing the internal and external energy functions using the generalized gradient vector flow field (GGVF) [63]. The detected envelopes are used to extract several indices (e.g., velocity time integral). These indices are strongly correlated (r=0.99) with the indices computed manually by human experts. Other contour-based method for cardiac and Doppler segmentation are presented in [91], [92]. As shown in these works, the shape and location of the initial contour greatly impacts the process of segmentation; additionally, these methods might require manual annotations and tend to be computationally expensive.

Model-Based Methods: Kalinic et al. [93] proposed a model-based method for segmenting the velocity profile in CW images. The segmentation process consists of two main steps: registration step and transferring step. The registration step generates a set of parameters that describe the geometric transformation of target-reference mapping. The reference image (model) is chosen to be the least different from the rest of the images in the dataset. This reference image is chosen by calculating the mutual mappings of all the images as described in [93]. After selecting the reference image, the velocity profile of this image is segmented manually by a cardiologist. The velocity profile of a new target image can be obtained by geometrically transferring, using the parameters obtained in the registration step, the profile of the reference image (model) to target image. The proposed method is evaluated using 59 velocity profiles extracted manually from CW images [93]. Instead of manually selecting the reference image, Kalinic et al. [94] extended [93] and used an atlas generated from CW Doppler images of healthy volunteers to register a new target image. The atlas image (reference image) is the statistical average of all images constructed using the arithmetic image averaging operation as discussed in [94]. Detailed presentation of the atlas model and discussion of results can be found in [94]. A recent model-based method that uses an atlas for constructing the spectral envelope is presented in [95].

Although model-based methods have been successfully used in spectral envelopes segmentation, these methods have difficulty handling the Doppler variations among patients and various disease types. Furthermore, they require manual annotation, and can become computationally expensive.

Machine Learning-Based Methods: Park et al. [96] introduced a learning-based method for detecting the spec-

tral envelope of mitral valve (MV) inflow. The method starts by training a series of detectors to detect a left root point (E velocity), right root point (A velocity), a single triangle box (E and A velocities overlapped), and a double triangle box (E and A velocities separated). Each of these detectors, which are trained using negative and positive examples, provided label and detection probability. After identifying the region of interest using these detectors, the triangle shape was inferred using a shape inference algorithm [96]. Given training images and their corresponding shapes, this algorithm learns a non-parametric regression function that gives a mapping from an image to its shape. Once the shape profiles are generated, the best shape among all candidates is selected as the final spectral envelope shape. Finally, four flow measurements [96] are computed from the detected envelopes. This method is evaluated using 298 Doppler images and compared with manually traced envelopes. The experimental results presented in [96] proved the superiority of the proposed method as compared to a previous method [97]. Known limitations of machine learning-based methods are: 1) the need of a large number of manually labeled images and 2) the subjectivity and difficulty of extracting the best set of features.

In summary, we present above several automated methods for spectral envelope segmentation. Existing methods for Doppler segmentation have different limitations that need to be addressed to obtain robust and practical automated clinical applications. For example, current methods are sensitive to the image noise, variations, and are designed for specific Doppler profile. Future research, therefore, should focus on developing segmentation methods that are robust to noise, different variations, and blood flows. Another future direction would be to automate Doppler gate localization to speed up the acquisition process, increase the quality of the recorded Doppler, and enhance the segmentation performance. Finally, an interesting direction for future research would be to investigate the use of recent deep learning methods for spectral envelope segmentation.

In addition to the aforementioned methods, we refer the reader to automated non-image methods applied directly to the raw signal for maximum velocity estimation [98], [99], [100], [101]. These methods are highly affected by the signal-to-noise levels and the transducer configurations. Further, they can only be applied to the original Doppler signal during the acquisition.

4.4.3   M-mode, Wall Segmentation

M-mode echo is used to provide an accurate assessment of small cardiac structures with rapid motions (e.g., valves). Assessing cardiac function from M-mode requires to accurately delineate wall boundaries followed by estimating different indices (e.g., left ventricular dimension at end-systole). This process is challenging due to the presence of image artifacts and false echoes

between the cardiac walls. Contrary to B-mode and Doppler images, only few methods are proposed to delineate wall boundaries in M-mode images.

For example, Fancourt et al. [102] proposed a fully automated method for delineating anterior and posterior walls in M-mode images. The method starts by splitting an M-mode image into anterior wall and posterior wall regions. For each region, the relative distance offsets between all pairs of scans are calculated using cross-correlation [102]. These offsets are converted to relative wall motion using global optimization followed by calculating the absolute wall motion from the relative wall motion using interpolation (interpolating over M-mode images). The proposed method is designed and evaluated using a small and invariant dataset. In summary, very few automated methods are proposed to segment cardiac walls in M-mode images. Therefore, an important future direction would be to develop automated M-mode analysis methods using large datasets collected by different vendor/software from different populations.

4.5   Cardiac Disease Classification

The majority of automated echo methods for CVD classification focus on 1) detecting diseases that cause Wall Motion Abnormalities (WMA) based on analysis of B-mode or 2) evaluating cardiac dysfunction from Spectral Doppler. We are not aware of any automated method that uses M-mode for CVD classification.

4.5.1   CVD Classification from B-mode Echo

B-mode echo is commonly used for detecting and assessing wall motion abnormalities (WMA). These abnormalities are observed in several cardiac diseases such as cardiomyopathy and coronary artery disease (CAD) [2]. Four terms are usually used in echocardiography to describe different types of WMA: Hypokinetic (reduced movement), akinetic (lack of movement), dyskinetic (abnormal movement), and aneurysm (abnormal wideness) [2]. Cardiomyopathy is a disease of the heart muscle that can cause abnormal dilation, thickening or lack of function of focal segments of the heart. Dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and ischemic cardiomyopathy (IC) are three major cardiomyopathy diseases. DCM is a cardiac muscle disease that enlarges LV wall and causes abnormal global motion. HCM is another muscle disease that causes thickening of the cardiac muscle (myocardium), which can lead to stiffness of LV as well as global and regional motion abnormalities. Ischemic cardiomyopathy (IC) causes weakness of the cardiac muscles [2]. CAD (coronary artery disease) occurs when the coronary arteries become narrowed or blocked. MI (myocardial infarction) is a serious cardiac disease that occurs as a result of severely narrowed or blocked coronary artery. Coronary artery disease can be detected

by the presence of regional WMA on echocardiogram [2].

Several machine learning-based methods published in the literature can detect WMA, CAD, and cardiomyopathy diseases based on automatically extracted B-mode indices (e.g., LV volume) or disease-relevant features extracted directly from the images.

For example, Leung and Bosch [103] proposed an automated method to assess WMA. The proposed method is developed and evaluated using B-mode data (A2C and A4C) collected from 129 random patients; data of 65 patients are used for training and data of 64 patients are used for testing. The ground truth (LV endocardial contours) is provided using a semi-automated technique and further validated by two cardiologists. Scores of abnormalities are also provided by the cardiologists as follows: 0 = normokinesia, 1 = hypokinesia, 2 = akinesia, and 3 = dyskinesia. These scores are grouped to create two classes: normal motion (score of 0) and abnormal motion (score > 0). The annotated contours are used to construct LV shape model, which is further analyzed using PCA to extract statistical parameters for abnormality classification. Different combination of PCA shape modes and parameters are used to train the classifier. In all cases, higher correct classification rate is achieved using less shape parameters. The trained binary classifier achieved up to 91.1% average accuracy in classifying wall motions as normal or abnormal. Similarly, Qazi et al. [104] used a shape-based method to automatically delineate the boundary of LV in each frame. Then, several cardiac structural and functional features, namely circumferential and radial strains, as well as local, segmental, and global Simpson volumes, are extracted from the delineated LV shape. The extracted numerical features are then reduced (Kolmogorov-Smirnov test) to select the best features for training the classifier. The trained classifier, tested using 220 cases, achieved a sensitivity that ranges from 80% to 90% in classifying cases as normal or abnormal (hypokinetic, akinetic, dyskinetic, and aneurysm).

Shalbaf et al. [105] proposed quantitative regional index for WMA detection and CAD predication. The proposed method is evaluated using 345 cases (B-mode, A2C and A4C) collected from 10 healthy volunteers and 14 patients with CAD. The ground truth labels, which include LV region, landmarks, and scores of abnormalities, are annotated by a group of trained cardiologists. The proposed method combines affine transformation and B-spline snake to delineate LV and calculate a novel index for WMA classification. Specifically, the proposed index is computed from the control points of B-spline snake model. For classification, two threshold values, determined using the quantitative regional indices of all images in the training set, are used. The determined thresholds are used to classify the testing set (125 cases) as normal or abnormal (hypokinetic, akinetic, dyskinetic, aneurysm). The agreement between the scoring of abnormalities obtained by the proposed index and those assigned by two experts achieved 83% absolute agreement and 99% relative agreement.

For CAD risk assessment, Araki et al. [106] introduces a method for classifying patients as high or low risk. The method starts by extracting 56 types of grayscale features that represent the coronary texture directly from the image. Examples of these features include gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), intensity histogram, gray level difference statistics (GLDS), neighborhood gray tone difference matrix (NGTDM), invariant moment, and statistical feature matrix (SFM). Then, six combinations of features are generated, and the best combination is chosen based on classification accuracy. The best set is used to train Support vector machine (SVM) for CAD risk assessment. The method is evaluated using 2865 B-mode frames collected from 15 patients. These frames are labeled, using stroke-risk biomarker ($cIMT > 0.9$mm), as high-risk (1508) and low-risk (1357). To select the best kernel of SVM and the best set of feature combination, K-fold cross validation protocol with 10 trials is used. The proposed method achieved up to 94.95% average accuracy and 0.95 AUC in classifying patients as low-risk and high-risk. Other machine learning methods for CAD detection and risk assessment can be found [107] (first-order statistical features, ANOVA for reduction, and NN classifier) , [108] (trace transform and fuzzy texture), [109] (discrete wavelet transform and marginal fisher analysis), and [110] (GLCM and SVM).

Sudarshan et al. [111] presented a machine learning framework for myocardial infarction (MI) detection and assessment. For feature extraction, Local Configuration Pattern (LCP) descriptor is used to extract 17850 LCP features from 46200 Curvelet Transform (CT) coefficients of echo. Prior to classification, the extracted features are reduced using Marginal Fisher Analysis (MFA) followed by fuzzy entropy based ranking method (mRMR) to select the best set of features. The proposed framework achieved an accuracy of 98.99%, sensitivity of 98.48%, and specificity of 100% using Support Vector Machine (SVM) classifier with only six features. In addition to handcrafted features, a novel index, called Myocardial Infarction Risk Index (MIRI), is proposed to detect three types of MI: normal, moderate, and severe. MIRI is generated by combining the most distinguishing features of MFA, and it is formulated as follows: MIRI = (0.15 × MFA 8) + (0.3 × MFA 1) + 2.5. The mean values of MIRI for normal, moderate and severely MI are 6.6, 7.4, and 5.9, respectively. Using the proposed index for the identification of MI stages achieved excellent performance comparable to the performance of classification using handcrafted (LCP) features. We refer the reader to other automated method for MI detection and assessment [112] (DWT, GLCM, and higher-order moment spectra [HOS] features and

SVM) and [113] (HOS, Fractal Dimension (FD), Hu moments, Gabor features and SVM).

Automated methods for detecting and diagnosing dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM) are proposed in [114] and [115]. In [114], the automated method starts by denoising each frame followed by segmenting LV in that frame using Fuzzy c-means (FCM). The segmented LV is used to extract cardiac parameters such as volume and ejection fraction (EF). In addition to these parameters, principal component analysis (PCA) and discrete cosine transform (DCT) algorithms are applied to the segmented LV to extract shape and statistical features for DCM and HCM diagnosis. The extracted PCA and DCT features are used with NN, SVM and combined K-NN to detect normal hearts, hearts affected with DCM, and hearts affected with HCM. The experimental results showed that the highest performance (92.04%) in classifying normal and affected heart is obtained using PCA features with NN classifier. It also showed that PCA features are better than DCT and cardiac indices (e.g., ejection fraction) for DCM and HCM diagnosis because moderately and mildly abnormal cases can have normal indices values.

Narula et al. [115] used an ensemble of three machine learning classifiers, namely SVM, random forests (RF), and neural networks (NN), to automatically differentiate between hypertrophic cardiomyopathy (HCM) and physiological hypertrophy in athletes (ATH). The proposed ensemble approach is developed and evaluated using data obtained from 77 ATH and 62 HCM patients. Several geometric (e.g., LV diameter) and mechanical (e.g., strain) indices are extracted from the delineated chamber using a commercial software, and further reduced using information gain (IG) algorithm. The output of IG algorithm revealed that volume (IG = 0.24), mid-left ventricular segmental (IG = 0.134), and average longitudinal strain (IG = 0.131) are the best features or predictors for differentiating between HCM and ATH. The model, which was evaluated using 10-fold cross validation, achieved 87% sensitivity and 82% specificity in distinguishing HCM from ATH. It achieved 96% sensitivity when adjusted for age. The paper concluded the capability of machine-learning algorithms to accurately discriminate between physiological and pathological patterns of hypertrophic remodeling.

4.5.2 CVD Classification from Doppler Echo

Doppler outperforms other imaging modalities in assessing valve regurgitation and stenosis [116]. The accurate detection of valve dysfunction relies heavily on the Doppler indices extracted from the spectral envelopes (Section 4.4.2).

Kiruthika et al. [83] proposed an automated method for assessing the severity of aortic valve regurgitation. To delineate the spectral envelopes and extract three Doppler indices, several low level image processing-based techniques, namely filtering, morphological oper-

ations, thresholding, and edge detection, are used. Once the spectral envelope is delineated, 3 Doppler indices are extracted: peak pressure gradient (PPG), peak flow velocity (PFV), and pressure half time (PHT). These indices showed strong correlation with manual indices when applied to 22 images of 11 patients with mild, moderate, and severe aortic regurgitation; i.e., assessment of aortic regurgitation severity showed a strong positive correlation (r=0.95).

Another method for quantifying the severity of valve dysfunctions is presented in [63]. Deformable-based method (active contour) is used to delineate the spectral envelope of the left ventricular outflow tract (LOVT) and transvalvular flow (TF) in 30 patients with aortic or mitral stenosis, 20 with normal sinus rhythm and 10 with atrial fibrillation. The delineated envelopes are then used to extract three important Doppler indices: the maximum velocity (V_{max}), the mean velocity (V_{mean}), and the velocity time integral (VTI). Comparison between the automatically extracted indices and manual indices extracted by two experienced echocardiographers showed good agreement. In addition, the results of B&A analysis on V_{max}, V_{mean}, and VTI for all patients showed acceptable limits of agreement and small bias; -3.9% to +0.5% (V_{max}), -4.6% to -1.4% (V_{mean}), and -3.6% to +4.4% (VTI).

Kalinic et al. [94] presented a method for detecting aortic stenosis (AS) and coronary artery disease (CAD) using Doppler indices extracted from spectral envelopes. The segmentation of the spectral envelope is performed by registering the input image to an atlas. The registration step consists of geometric transformation, similarity measure, and optimization. The atlas, which is used as a template for segmentation, is constructed from spectral envelopes of healthy volunteers (59 envelopes). The proposed method is validated using 36 profiles belong to patients with CAD, 53 profiles belong to patients with AS, and 59 profiles belong to healthy volunteers. Once the envelope is segmented, three Doppler indices are extracted and used for disease detection. These indices are time-to-peak, peak value, and rise–fall time ratio. The experimental results showed strong statistical correlation between the parameters extracted automatically and those extracted manually by the expert cardiologist.

Similar to [94], disease-specific atlases are constructed using a proposed hybrid framework [95]. Specifically, two atlases are created from the aortic Doppler images of 100 healthy individuals and 100 patients with AS and used as template for segmentation. After segmenting the envelope, four diagnostic values, namely area, jet velocity, mean and peak gradient, are extracted and combined with physiological parameters (e.g., heart rate) to detect 3 levels of AS: mild, moderate, and severe. The experimental results showed comparable segmentation and assessment performance between the automated and manual methods (see [95]

for results).

Table 5 provides a summary of automated methods for CVD classification from B-mode and Doppler echo. We refer the reader to [106], [117] for comprehensive discussions of CAD, MI, and HCM diseases. Further, comprehensive discussions of Doppler disease-specific features and the importance of automated diagnosis of valve regurgitation and stenosis can be found in [118], [119]. In summary, automated CVD classification has attracted researchers and clinicians in the past decades. Most existing works, however, focus on detecting LV dysfunction or diagnosing its abnormality. Future research should focus on analyzing other chambers (e.g., RV and LA) as well as all chambers together (whole heart). Existing works also focus on analyzing normal or slightly abnormal cardiac structures, and predicting common cardiac diseases (e.g., DCM). Therefore, future research should focus on analyzing abnormal structures and rare diseases. One obvious limitation of this direction is the lack of datasets acquired from patients with abnormal structures and rare diseases. Finally, future works should focus on developing automated CVD classification for fetuses and neonates because less attention has been paid to these populations as compared to adults.

5 ECHOCARDIOGRAPHY DATASETS

As the performance of automated echo analysis system depends highly on the data, there is a critical need for collecting well-annotated, diverse, and relatively large echo datasets. Several datasets are publicly available for cardiac CT and MRI modalities. Examples of these datasets include MESA [120], Cardiac MRI [121], SCD [122], DETERMINE [123], SCMR [124], and CHD [125] for CMR modality, and Left atrial CT dataset [126] for CT modality. In case of echocardiography, we are not aware of any Doppler or M-mode datasets that are publicly available for research use. We are aware of only three B-mode echo datasets made available publicly for researchers. These datasets are EchoNet-Dynamic [127], CETUS [128], and STACOM [129]. The characteristics of these datasets are presented in Table 6.

5.1 EchoNet-Dynamic Dataset

EchoNet-Dynamic dataset [127] contains 10,025 echo videos (2D B-mode, A4C) collected from 10,025 patients admitted to Stanford University Hospital between 2006 and 2018. Patients average age is 68 ± 21 and 49% of them are female. The number of patients in training, validation, and testing sets are 7460, 1288, and 1277, respectively. Videos of the dataset are recorded from different angles, locations, and image acquisition techniques (e.g., iE33, Sonos, Acuson SC2000, Epiq 5G). Each video is de-identified and cropped to get the anatomical region. The cropped region ($600x \times 600$ or 768×768) is

then downsampled, using cubic interpolation, into standardized 112×112 pixels. In each video, LV boundary is traced at end-systole (ES) and end-diastole (ED) frames by a human expert. In addition to the videos, the dataset contains demographic information and cardiac indices obtained by a registered sonographer and verified by an echocardiographer. These indices are ejection fraction (EF), end-systolic volume (ESV), and end-diastolic volume (EDV). To the best of our knowledge, EchoNet-Dynamic is the largest labeled echo dataset made available publicly (via Github) to the research community.

5.2 CETUS Dataset

CETUS [128] is another open access 3D-US (B-mode, A4C) dataset for the automatic delineation of LV borders at ED and ES frames. It was released in conjunction with the Challenge on "Endocardial Three-dimensional Ultrasound Segmentation" during MICCAI 2014. The dataset has 3D-US images collected from 45 subjects, divided into three groups: 1) 15 healthy subjects, 2) 15 patients with a history of myocardial infarction (MI), and 3) 15 patients with a history of dilated cardiomyopathy (DCM). The data are acquired at three different hospitals: Rennes University Hospital-France, University Hospital Leuven-Belgium, and Erasmus MC-Netherlands. In addition, the data acquisition is performed using three different machines: a GE Vivid E9, using 4V probe, a Philips iE33, using either X3-1 or a X5-1 probe, and a Siemens SC2000, using 4Z1c probe. Using these machines, each hospital collected data for 5 subjects from each group. This data collection setup ensures that the group, hospital, and ultrasound machines are equally distributed. Images are excluded from further analysis if they violate a set of pre-determined criteria (see Table 6). The dataset ground truth segmentations are obtained by 3 expert cardiologists using a non-commercial contouring package (Speqle3D) at ED and ES frames. The data and their corresponding ground truth segmentation are divided into a training set (15 subjects) and a testing set (30 subjects).

5.3 STACOM Dataset

STACOM [129] is an open access benchmark that was prepared for a challenge called "Cardiac Motion Analysis Challenge" at MICCAI 2012. The benchmark includes data of two modalities (MRI and 3D-US) and 16 anatomies, each has 13 to 30 frames. The data of 15 healthy volunteers (aged 28 ± 5 years, 19% female) without a clinical history of cardiac disease and one dynamic phantom are recruited at the Division of Imaging Sciences and Biomedical Engineering (King's College London, UK) and the Department of Internal Medicine (University of Ulm, Germany), respectively. The ultrasound data are acquired using iE33 3D echocardiography system with a $3D \times 301$ matrix array transducer in full-volume acquisition (FVA) mode. All the data

are acquired from the left ventricle (LV) in apical view during breath-hold to minimize artifacts. A total of 12 landmarks (4 walls at 3 ventricular levels) was manually tracked by two observers over the whole cardiac cycle using an in-house application. These points were registered to 3D coordinates using a point based similarity transform. The median of inter-observer variability for phantom and volunteer datasets are 0.77 mm and 0.84 mm, respectively.

The benchmarks (data and ground truth) for CE-TUS and STACOM datasets are publicly available for research use via Cardiac Atlas Project.

6 CURRENT LIMITATIONS AND FUTURE DIRECTIONS

Despite the high imaging quality of CMR and CT, echo remains the most popular and commonly used modality for diagnosing CVD. This is mainly attributed to the portability, availability, less complexity, and lower cost of echo as compared to other modalities. These attributes make it possible, especially in low-resource settings, to take full advantage of echo and use it for diagnosis. However, the interpretation of the acquired echo data requires echocardiographic expertise which is lacking in low-resource settings. In addition, the manual interpretation is error-prone and suffers from intra-/inter-reader variability. In such cases, fully automated screening and diagnostic systems have a significant potential in mitigating subjectivity and providing high-quality and cost-efficient healthcare, especially for patients in low-resource settings.

In this paper, we reviewed existing automated methods for performing different echo tasks. These methods achieved good to excellent performance and proved the feasibility of using fully automated systems for acquisition, interpretation, and diagnosis. Therefore, the question arises of whether or not automated echo screening and diagnosis systems are ready to be incorporated into the clinical practice. Our extensive review revealed that several issues and limitations need to be addressed prior to using fully automated systems in clinical practice, point-of-care ultrasound (POCUS), and low-resource settings.

Performance of automated systems: how much accuracy is acceptable or enough? Current methods proved the feasibility of using automated echo interpretation and diagnosis. However, it is not clear if the accuracy of these methods is acceptable for diagnosis in clinical practice; i.e., the impact of the obtained accuracy on clinical outcomes is not clear and requires further investigation. Thus, it is important for future research to focus attention on not only the technical development but also on measuring the quality of automated methods (e.g., [130]) and the actual impact of these methods on clinical outcomes.

Similar acquisition configurations and unrepresentative datasets. Most existing methods are designed using datasets collected by specific devices under specific configurations; i.e., they are sensitive to the acquisition's devices and configurations. Therefore, we believe another important future direction would be to provide a systematic comparison of existing devices/configurations for echo acquisition and study their impact on performance. In addition, we believe future methods should focus on using datasets collected by various devices under different configurations to enhance generalizability. Another major limitation of the most current studies is that they are built and evaluated using relatively small and invariant datasets. This can lead to significant variations in performance across different datasets. Hence, future research should also focus on using relatively large and variant datasets for developing robust systems. The use of large and representative datasets is especially important when developing deep learning-based methods.

Limited automated echo acquisition methods. The majority of automated methods are applied to echo images after acquisition to perform view classification, quality assessment, region segmentation, indices calculation, or CVD diagnosis. To speed up the acquisition process, future research should focus on automating echo screening and acquisition tasks. For example, locating the optimal imaging plane and sample volume (gate) requires time and expertise. Therefore, developing automated gate localization methods can decrease the acquisition complexity, reduce the time of manual gate adjustments, and increase reproducibility.

Limited methods for M-mode and Color Doppler. Among all echo modes, B-mode, especially the apical, short, and long axis views, received the most attention followed by spectral Doppler. Only few automated methods are proposed to analyze M-mode, color Doppler, and rare B-mode views. M-mode images is commonly used to diagnose several cardiac diseases in fetuses. Color Doppler is well-suited for assessing valves regurgitation and stenosis as well as detecting septal defects and intracardiac shunts. Future research, therefore, should focus on developing robust systems that can interpret all views and modes, including M-mode and color Doppler.

LV chamber analysis is the primary focus of most existing automated methods. As LV chamber plays a critical role in blood circulation and the diagnosis of several CVD, existing methods focus mainly on segmenting and analyzing this chamber. RV, LA, and RA chambers have received less attention due to their complex shape and unclear boundaries. Because cardiac indices extracted from RA, LA, and RV chambers are also important for diagnosing various CVD, future works should develop fully automated methods that can handle the complex structures of these chambers. Future research should

also focus on segmenting the heart as a whole to enable global assessment that considers the combined motion of all chambers together.

Population-specific methods evaluated using normal and homogeneous cases. Existing automated methods paid less attention to fetuses/neonates/children, and focus mainly on adults. Because cardiovascular systems of adults and fetal exhibit significant differences [131], these methods would perform poorly or might completely fail when applied to fetuses/neonates/children. In addition, automated methods that are developed and evaluated using normal datasets might not work on cases that have abnormal or greater than mild pathological deformities. To address these issues, future works should focus on 1) developing cross-population methods or 2) population-specific methods for fetal, neonatal, and adult. Future works should also focus on developing algorithms robust to image inhomogeneity, pathological deformities, and shape irregularities, to assist in analyzing and predicting rare CVD.

Learning multiple echocardiography tasks in isolation. Automated echo image analysis typically includes several tasks such as noise reduction, detection, segmentation, and classification. These tasks are often implemented through separate machine learning methods. This approach of analysis involves unnecessary repetitions and limits the actual impact of machine learning. Therefore, it is important to use advanced machine learning methods, such as multitask learning, to simultaneously learn several related echo tasks at once. This approach would improve generalization and decreases resource utilization while preventing unnecessary repetitions of building task-specific models in isolation.

Lack of interpretable and explainable automated echo methods. Current automated echo methods generate an output, based on features extracted from the images, without providing feature importance weights or an explanation for the detected output. The lack of explainability and transparency can lead to unreliable decision making (black-box decisions). Hence, future work should focus on integrating explainability into automated systems using approaches that range from Global variable importance measures to ICNN (interpretable convolutional neural networks) [132] or model-agnostic explanations [133].

Lack of automated scientific discoveries in cardiology. current methods are designed, using labeled training data, to compute established parameters (e.g., LV volume). These parameters are then fed into machine learning classifiers to detect known patterns. Supervised learning approach relies on expert's knowledge, and therefore, cannot extract knowledge unknown to the experts. To automate scientific discoveries in cardiology, it is important to explore unsupervised and reinforcement learning approaches. These approaches can detect new patterns, extract knowledge unknown to experts, explore different actions, and learn which actions lead to a better diagnosis.

Current echo systems are partially automated. Current automated systems require experts to manually localize Doppler gate, detect ROIs, or select ED/ES frames prior to performing a specific cardiac image analysis task (e.g., segmentation). Future research, therefore, should focus on developing fully automated (end-to-end) systems that could be used efficiently in real-time to acquire data, analyze desired views and frames, segment cardiac region, and diagnose diseases without, or with minimum, user intervention. Such fully automated systems have the potential of reducing clinical workflow as well as improving patient outcomes. These systems can be used in POCUS and low-resource settings to provide high-quality and cost-efficient healthcare.

Automated real-time echo analysis. Several challenges need to be addressed before adopting automated real-time echo analysis and integrating it into clinical and low-resource settings. For example, existing methods focus mainly on improving performance. Little consideration has been given to issues such as speed, computation time, memory usage, model size, power/energy consumption, and scalability. Future works, therefore, should consider these issues and design lightweight systems that can achieve maximum performance in latency-sensitive applications and resource-limited environments.

Dataset and code availability. Open-access code and publicly available datasets can speed up and strengthen advances in automated echo analysis because it facilitates reproducibility of results and allows to extend existing methods. Hence, it is very valuable for future works to provide the data and code necessary for replication and improvement. So far, we are aware of only few echo datasets and codes [46], [127] that are publicly available for research use. We described these publicly available echo datasets in Section 5.

7 CONCLUSION

Automated echo analysis is critical to improve the limitations of current practice and provide high-quality healthcare to the patients in low-resource settings. The first step of any automated system involves accurately detecting ROI as it highly impacts the performance of subsequent automated tasks (e.g., segmentation). In this paper, we reviewed automated ROI detection methods as well as automated methods of four cardiac tasks: echo quality assessment, mode/view classification, boundary segmentation, and CVD diagnosis. We also provided a summary of publicly available echo datasets followed by a thorough discussion of current limitations and potential future directions. This paper provides biomedical engineers and clinicians a stand-alone summary of automated echocardiography analysis and interpretation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

18

ACKNOWLEDGEMENT

REFERENCES

[1] E. J. Benjamin, P. Muntner, and M. S. Bittencourt, "Heart disease and stroke statistics-2019 update: A report from the american heart association," *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019.

[2] C. Mitchell, P. S. Rahko, L. A. Blauwet, B. Canaday, J. A. Finstuen, M. C. Foster, K. Horton, K. O. Ogunyankin, R. A. Palma, and E. J. Velazquez, "Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: Recommendations from the american society of echocardiography," *Journal of the American Society of Echocardiography*, vol. 32, no. 1, pp. 1–64, 2019.

[3] C. C. Nwabuo, H. T. Moreira, H. D. Vasconcellos, G. Yared, B. A. Venkatesh, J. P. Reis, P. J. Schreiner, D. M. Lloyd-Jones, K. Ogunyankin, C. E. Lewis, *et al.*, "Inter-and intra-reader reproducibility of left ventricular volumetric and deformational assessment by three-dimensional echocardiography in a multi-center community-based study: The coronary artery risk development in young adults (cardia) study," 2016.

[4] M. G. Khouri, B. Ky, G. Dunn, T. Plappert, V. Englefield, D. Rabineau, E. Yow, H. X. Barnhart, M. S. J. Sutton, and P. S. Douglas, "Echocardiography core laboratory reproducibility of cardiac safety assessments in cardio-oncology," *Journal of the American Society of Echocardiography*, vol. 31, no. 3, pp. 361–371, 2018.

[5] M. S. Chew and J. Poelaert, "Accuracy and repeatability of pediatric cardiac output measurement using doppler: 20-year review of the literature," *Intensive care medicine*, vol. 29, no. 11, pp. 1889–1894, 2003.

[6] V. Chetboul, D. Concordet, J. Pouchelon, N. Athanassiadis, C. Muller, L. Benigni, A. Munari, and H. Lefebvre, "Effects of inter-and intra-observer variability on echocardiographic measurements in awake cats," *Journal of Veterinary Medicine Series A*, vol. 50, no. 6, pp. 326–331, 2003.

[7] Z. Ahmed, M. Saada, A. M. Jones, and A. M. Al-Hamid, "Medical errors," 2019.

[8] G. Litjens, F. Ciompi, J. M. Wolterink, B. D. de Vos, T. Leiner, J. Teuwen, and I. Išgum, "State-of-the-art deep learning in cardiovascular image analysis," *JACC: Cardiovascular Imaging*, vol. 12, no. 8, pp. 1549–1565, 2019.

[9] K. M. Meiburger, U. R. Acharya, and F. Molinari, "Automated localization and segmentation techniques for b-mode ultrasound images: A review," *Computers in biology and medicine*, vol. 92, pp. 210–235, 2018.

[10] D. Kang, J. Woo, C. J. Kuo, P. J. Slomka, D. Dey, and G. Germano, "Heart chambers and whole heart segmentation techniques," *Journal of Electronic Imaging*, vol. 21, no. 1, p. 010901, 2012.

[11] S. Mazaheri, P. S. B. Sulaiman, R. Wirza, F. Khalid, S. Kadiman, M. Z. Dimon, and R. M. Tayebi, "Echocardiography image segmentation: A survey," in *2013 International Conference on Advanced Computer Science Applications and Technologies*, pp. 327–332, IEEE, 2013.

[12] P. Peng, K. Lekadir, A. Gooya, L. Shao, S. E. Petersen, and A. F. Frangi, "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 29, no. 2, pp. 155–195, 2016.

[13] D. E. O. Dewi, H. N. Abduljabbar, and E. Supriyanto, "Review on advanced techniques in 2-d fetal echocardiography: an image processing perspective," in *Advances in Medical Diagnostic Technology*, pp. 53–74, Springer, 2014.

[14] M. Alsharqi, W. Woodward, J. Mumith, D. Markham, R. Upton, and P. Leeson, "Artificial intelligence and echocardiography," *Echo research and practice*, vol. 5, no. 4, pp. R115–R125, 2018.

[15] V. Sudarshan, U. R. Acharya, E. Y.-K. Ng, C. S. Meng, R. San Tan, and D. N. Ghista, "Automated identification of infarcted myocardium tissue characterization using ultrasound images: a review," *IEEE reviews in biomedical engineering*, vol. 8, pp. 86–97, 2014.

[16] M. K. Feldman, S. Katyal, and M. S. Blackwood, "Us artifacts," *Radiographics*, vol. 29, no. 4, pp. 1179–1189, 2009.

[17] T. Joel and R. Sivakumar, "Despeckling of ultrasound medical images: A survey," *Journal of Image and Graphics*, vol. 1, no. 3, pp. 161–165, 2013.

[18] F. Benzarti and H. Amiri, "Speckle noise reduction in medical ultrasound images," *arXiv preprint arXiv:1305.1344*, 2013.

[19] A. Perperidis, "Postprocessing approaches for the improvement of cardiac ultrasound b-mode images: A review," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 3, pp. 470–485, 2016.

[20] G. N. H. A. El-Gwad and Y. M. Omar, "Selection of the best despeckle filter of ultrasound images," in *2017 2nd International Conference on Multimedia and Image Processing (ICMIP)*, pp. 245–249, IEEE, 2017.

[21] P. Wang, H. Zhang, and V. M. Patel, "Sar image despeckling using a convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1763–1767, 2017.

[22] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[23] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[24] V. Yeghiazaryan and I. Voiculescu, "An overview of current evaluation methods used in medical image segmentation," *Department of Computer Science, University of Oxford*, 2015.

[25] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, pp. 1–4, Springer, 2009.

[26] D. Giavarina, "Understanding bland altman analysis," *Biochemia medica: Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.

[27] S. R. Snare, H. Torp, F. Orderud, and B. O. Haugen, "Real-time scan assistant for echocardiography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 59, no. 3, pp. 583–589, 2012.

[28] S.-K. Pavani, N. Subramanian, M. D. Gupta, P. Annangi, S. C. Govind, and B. Young, "Quality metric for parasternal long axis b-mode echocardiograms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 478–485, Springer, 2012.

[29] A. H. Abdi, C. Luong, T. Tsang, J. Jue, K. Gin, D. Yeung, D. Hawley, R. Rohling, and P. Abolmaesumi, "Quality assessment of echocardiographic cine using recurrent neural networks: Feasibility on five standard view planes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 302–310, Springer, 2017.

[30] A. H. Abdi, C. Luong, T. Tsang, G. Allan, S. Nouranian, J. Jue, D. Hawley, S. Fleming, K. Gin, J. Swift, *et al.*, "Automatic quality assessment of apical four-chamber echocardiograms using deep convolutional neural networks," in *Medical Imaging 2017: Image Processing*, vol. 10133, p. 101330S, International Society for Optics and Photonics, 2017.

[31] G. Zamzmi, L.-Y. Hsu, W. Li, V. Sachdev, and S. Antani, "Echo doppler flow classification and goodness assessment with convolutional neural networks," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 1744–1749, IEEE, 2019.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

19

[32] H. Wu, D. M. Bowers, T. T. Huynh, and R. Souvenir, "Echocardiogram view classification using low-level features," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, pp. 752–755, IEEE, 2013.

[33] Y. Qian, L. Wang, C. Wang, and X. Gao, "The synergy of 3d sift and sparse codes for classification of viewpoints from echocardiogram videos," in *MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support*, pp. 68–79, Springer, 2012.

[34] D. Agarwal, K. Shriram, and N. Subramanian, "Automatic view classification of echocardiograms using histogram of oriented gradients," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, pp. 1368–1371, IEEE, 2013.

[35] J. H. Park, S. K. Zhou, C. Simopoulos, J. Otsuki, and D. Comaniciu, "Automatic cardiac view classification of echocardiogram," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[36] S. Ebadollahi, S.-F. Chang, and H. Wu, "Automatic view recognition in echocardiogram videos using parts-based representation," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, pp. II–II, IEEE, 2004.

[37] R. Kumar, F. Wang, D. Beymer, and T. Syeda-Mahmood, "Echocardiogram view classification using edge filtered scale-invariant motion features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 723–730, IEEE, 2009.

[38] O. A. Penatti, R. d. O. Werneck, W. R. de Almeida, B. V. Stein, D. V. Pazinato, P. R. M. Júnior, R. d. S. Torres, and A. Rocha, "Mid-level image representations for real-time heart view plane classification of echocardiograms," *Computers in biology and medicine*, vol. 66, pp. 66–81, 2015.

[39] M. Otey, J. Bi, S. Krishna, B. Rao, J. Stoeckel, A. Katz, J. Han, and S. Parthasarathy, "Automatic view recognition for cardiac ultrasound images," in *International Workshop on Computer Vision for Intravascular and Intracardiac Imaging*, pp. 187–194, 2006.

[40] S. K. Zhou, J. Park, B. Georgescu, D. Comaniciu, C. Simopoulos, and J. Otsuki, "Image-based multiclass boosting and echocardiographic view classification," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1559–1565, IEEE, 2006.

[41] S. R. Snare, S. A. Aase, O. C. Mjølstad, H. Dalen, F. Orderud, and H. Torp, "Automatic real-time view detection," in *2009 IEEE International Ultrasonics Symposium*, pp. 2304–2307, IEEE, 2009.

[42] G. Balaji, T. Subashini, and N. Chidambaram, "Automatic classification of cardiac views in echocardiogram using histogram and statistical features," *Procedia Computer Science*, vol. 46, pp. 1569–1576, 2015.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[44] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[46] J. Zhang, S. Gajjala, P. Agrawal, G. H. Tison, L. A. Hallock, L. Beussink-Nelson, E. Fan, M. A. Aras, C. Jordan, K. E. Fleischmann, *et al.*, "A computer vision pipeline for automated determination of cardiac structure and function and detection of disease by two-dimensional echocardiography," *arXiv preprint arXiv:1706.07342*, 2017.

[47] J. Zhang, S. Gajjala, P. Agrawal, G. H. Tison, L. A. Hallock, L. Beussink-Nelson, M. H. Lassen, E. Fan, M. A. Aras, C. Jordan, *et al.*, "Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy," *Circulation*, vol. 138, no. 16, pp. 1623–1635, 2018.

[48] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *NPJ digital medicine*, vol. 1, no. 1, p. 6, 2018.

[49] A. Østvik, E. Smistad, S. A. Aase, B. O. Haugen, and L. Lovstakken, "Real-time standard view classification in transthoracic echocardiography using convolutional neural networks," *Ultrasound in medicine & biology*, vol. 45, no. 2, pp. 374–384, 2019.

[50] H. Vaseli, Z. Liao, A. H. Abdi, H. Girgis, D. Behnami, C. Luong, F. T. Dezaki, N. Dhungel, R. Rohling, K. Gin, *et al.*, "Designing lightweight deep learning models for echocardiography view classification," in *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 10951, p. 109510F, International Society for Optics and Photonics, 2019.

[51] X. Gao, W. Li, M. Loomes, and L. Wang, "A fused deep learning architecture for viewpoint classification of echocardiography," *Information Fusion*, vol. 36, pp. 103–113, 2017.

[52] S. A. Melo, B. Macchiavello, M. M. Andrade, J. L. Carvalho, H. S. Carvalho, D. F. Vasconcelos, P. A. Berger, A. F. Da Rocha, and F. A. Nascimento, "Semi-automatic algorithm for construction of the left ventricular area variation curve over a complete cardiac cycle," *Biomedical engineering online*, vol. 9, no. 1, p. 5, 2010.

[53] J. C. Amorim, M. d. C. dos Reis, J. L. A. de Carvalho, A. F. da Rocha, and J. F. Camapum, "Improved segmentation of echocardiographic images using fusion of images from different cardiac cycles," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 511–514, IEEE, 2009.

[54] A. John and K. Jayanthi, "Extraction of cardiac chambers from echocardiographic images," in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, pp. 1231–1234, IEEE, 2014.

[55] J. Cheng, S. W. Foo, and S. Krishnan, "Automatic detection of region of interest and center point of left ventricle using watershed segmentation," in *2005 IEEE International Symposium on Circuits and Systems*, pp. 149–151, IEEE, 2005.

[56] J. Santos, D. Celorico, J. Varandas, and J. Dias, "Automatic segmentation of echocardiographic left ventricular images by windows adaptive thresholds," in *Proceedings of the International Congress on Ultrasonics, Vienna, April*, pp. 9–13, 2007.

[57] M. d. C. dos Reis, A. F. da Rocha, D. F. Vasconcelos, B. L. Espinoza, A. d. O. Francisco, J. L. de Carvalho, S. Salomoni, and J. F. Camapum, "Semi-automatic detection of the left ventricular border," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 218–221, IEEE, 2008.

[58] S. G. Lacerda, A. F. da Rocha, D. F. Vasconcelos, J. L. de Carvalho, I. G. Sene, and J. F. Camapum, "Left ventricle segmentation in echocardiography using a radial-search-based image processing algorithm," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 222–225, IEEE, 2008.

[59] F. A. A. Dawood, R. W. Rahmat, M. Z. Dimon, L. Nurliyana, and S. B. Kadiman, "Automatic boundary detection of wall motion in two-dimensional echocardiography images," *Journal of Computer Science*, vol. 7, no. 8, pp. 1261–1266, 2011.

[60] Y. Chen, F. Huang, H. D. Tagare, and M. Rao, "A coupled minimization problem for medical image segmentation with priors," *International journal of computer vision*, vol. 71, no. 3, pp. 259–272, 2007.

[61] A. Fernández-Caballero and J. M. Vega-Riesco, "Determining heart parameters through left ventricular automatic segmentation for heart disease diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2234–2249, 2009.

[62] M. Marsousi, A. Eftekhari, A. Kocharian, and J. Alirezaie, "Endocardial boundary extraction in left ventricular echocardiographic images using fast and adaptive b-spline snake algorithm," *International journal of computer assisted radiology and surgery*, vol. 5, no. 5, pp. 501–513, 2010.

[63] E. Gaillard, L. Kadem, P. Pibarot, and L.-G. Durand, "Optimization of doppler velocity echocardiographic measurements using an automatic contour detection method," in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

20

2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2264–2267, IEEE, 2009.

[64] M. Marsousi, J. Alirezaie, A. Ahmadian, and A. Kocharian, "Segmenting echocardiography images using b-spline snake and active ellipse model," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 3125–3128, IEEE, 2010.

[65] C. Tauber, H. Batatia, and A. Ayache, "A robust active contour initialization and gradient vector flow for ultrasound image segmentation.," in *MVA*, pp. 164–167, 2005.

[66] A. B. Oktay and Y. S. Akgul, "Echocardiographic contour extraction with local and global priors through boosting and level sets," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 46–51, IEEE, 2009.

[67] S. Nandagopalan, B. Adiga, C. Dhanalakshmi, and N. Deepak, "Automatic segmentation and ventricular border detection of 2d echocardiographic images combining k-means clustering and active contour model," in *2010 Second International Conference on Computer and Network Technology*, pp. 447–451, IEEE, 2010.

[68] M. Alessandrini, T. Dietenbeck, D. Barbosa, J. D'hooge, O. Basset, N. Speciale, D. Friboulet, and O. Bernard, "Segmentation of the full myocardium in echocardiography using constrained level-sets," in *2010 Computing in Cardiology*, pp. 409–412, IEEE, 2010.

[69] S. G. Antunes, J. S. Silva, J. B. Santos, P. Martins, and E. Castela, "Phase symmetry approach applied to children heart chambers segmentation: a comparative study," *IEEE transactions on biomedical engineering*, vol. 58, no. 8, pp. 2264–2271, 2011.

[70] A. Belaid, D. Boukerroui, Y. Maingourd, and J.-F. Lerallut, "Phase-based level set segmentation of ultrasound images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 138–147, 2011.

[71] W.-Y. Hsu, "Automatic atrium contour tracking in ultrasound imaging," *Integrated Computer-Aided Engineering*, vol. 23, no. 4, pp. 401–411, 2016.

[72] Y. Guo, Y. Wang, S. Nie, J. Yu, and P. Chen, "Automatic segmentation of a fetal echocardiogram using modified active appearance models and sparse representation," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1121–1133, 2014.

[73] D. Beymer, T. Syeda-Mahmood, A. Amir, F. Wang, and S. Adelman, "Automatic estimation of left ventricular dysfunction from echocardiogram videos," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 164–171, IEEE, 2009.

[74] G. Belous, A. Busch, and D. Rowlands, "Segmentation of the left ventricle from ultrasound using random forest with active shape model," in *2013 1st International Conference on Artificial Intelligence, Modelling and Simulation*, pp. 315–319, IEEE, 2013.

[75] J. Hansegard, S. Urheim, K. Lunde, and S. I. Rabben, "Constrained active appearance models for segmentation of triplane echocardiograms," *IEEE transactions on medical imaging*, vol. 26, no. 10, pp. 1391–1400, 2007.

[76] Y. Guo, L. Yu, Y. Wang, J. Yu, G. Zhou, and P. Chen, "Adaptive group sparse representation in fetal echocardiogram segmentation," *Neurocomputing*, vol. 240, pp. 59–69, 2017.

[77] B. Georgescu, X. S. Zhou, D. Comaniciu, and A. Gupta, "Database-guided segmentation of anatomical structures with complex appearance," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 429–436, IEEE, 2005.

[78] S. K. Zhou, "Shape regression machine and efficient segmentation of left ventricle endocardium from 2d b-mode echocardiogram," *Medical image analysis*, vol. 14, no. 4, pp. 563–581, 2010.

[79] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[80] Z. Yue, L. Wenqiang, J. Jing, J. Yu, S. Yi, and W. Yan, "Automatic segmentation of the epicardium and endocardium using convolutional neural network," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 44–48, IEEE, 2016.

[81] J. F. Silva, J. M. Silva, A. Guerra, S. Matos, and C. Costa, "Ejection fraction classification in transthoracic echocardiography using a deep learning approach," in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 123–128, IEEE, 2018.

[82] M. Zolgharni, N. M. Dhutia, G. D. Cole, K. Willson, and D. P. Francis, "Feasibility of using a reliable automated doppler flow velocity measurements for research and clinical practices," in *Medical Imaging 2014: Ultrasonic Imaging and Tomography*, vol. 9040, p. 90401D, International Society for Optics and Photonics, 2014.

[83] N. Kiruthika, B. Prabhakar, and M. R. Reddy, "Automated assessment of aortic regurgitation using 2d doppler echocardiogram," in *Proceedings of the 2006 IEEE International Workshop on Imagining Systems and Techniques (IST 2006)*, pp. 95–99, IEEE, 2006.

[84] N. Biradar, M. Dewal, and M. K. Rohit, "Automated delineation of doppler echocardiographic images using texture filters," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1903–1907, IEEE, 2015.

[85] A. Taebi, R. H. Sandler, B. Kakavand, and H. A. Mansy, "Estimating peak velocity profiles from doppler echocardiography using digital image processing," in *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–4, IEEE, 2018.

[86] N. M. Dhutia, M. Zolgharni, M. Mielewczik, M. Negoita, S. Sacchi, K. Manoharan, D. P. Francis, and G. D. Cole, "Open-source, vendor-independent, automated multi-beat tissue doppler echocardiography analysis," *The international journal of cardiovascular imaging*, vol. 33, no. 8, pp. 1135–1148, 2017.

[87] H. Kalinić, S. Lončarić, M. Čikeš, D. Miličić, and B. Bijnens, "Model-based segmentation of aortic ultrasound images," in *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 739–743, IEEE, 2011.

[88] M. Higa, P. Pilon, S. Lage, and M. Gutierrez, "A computational tool for quantitative assessment of peripheral arteries in ultrasound images," in *2009 36th Annual Computers in Cardiology Conference (CinC)*, pp. 41–44, IEEE, 2009.

[89] V. Magagnin, E. G. Caiani, L. Delfino, C. Champlon, S. Cerutti, and M. Turiel, "Semi-automated analysis of coronary flow doppler images: validation with manual tracings," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 719–722, IEEE, 2006.

[90] H. Greenspan, O. Shechner, M. Scheinowitz, and M. S. Feinberg, "Doppler echocardiography flow-velocity image analysis for patients with atrial fibrillation," *Ultrasound in medicine & biology*, vol. 31, no. 8, pp. 1031–1040, 2005.

[91] G. Zamzmi, L.-Y. Hsu, W. Li, V. Sachdev, and S. Antani, "Fully automated spectral envelope and peak velocity detection from doppler echocardiography images," in *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314, p. 113144G, International Society for Optics and Photonics, 2020.

[92] S. Zhu and R. Gao, "A novel generalized gradient vector flow snake model using minimal surface and component-normalized method for medical image segmentation," *Biomedical Signal Processing and Control*, vol. 26, pp. 1–10, 2016.

[93] H. Kalinić, S. Lončarić, M. Čikeš, D. Milicic, I. Čikeš, G. Sutherland, and B. Bijnens, "A method for registration and model-based segmentation of doppler ultrasound images," in *Medical Imaging 2009: Image Processing*, vol. 7259, p. 72590S, International Society for Optics and Photonics, 2009.

[94] H. Kalinić, S. Lončarić, M. Čikeš, D. Miličić, and B. Bijnens, "Image registration and atlas-based segmentation of cardiac

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

21

outflow velocity profiles," *Computer methods and programs in biomedicine*, vol. 106, no. 3, pp. 188–200, 2012.

[95] V. Baličević, H. Kalinić, S. Lončarić, M. Čikeš, and B. Bijnens, "A computational model-based approach for atlas construction of aortic doppler velocity profiles for segmentation purposes," *Biomedical Signal Processing and Control*, vol. 40, pp. 23–32, 2018.

[96] J. Park, S. K. Zhou, J. Jackson, and D. Comaniciu, "Automatic mitral valve inflow measurements from doppler echocardiography," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 983–990, Springer, 2008.

[97] S. K. Zhou, F. Guo, J. Park, G. Carneiro, J. Jackson, M. Brendel, C. Simopoulos, J. Otsuki, and D. Comaniciu, "A probabilistic, hierarchical, and discriminant framework for rapid and accurate detection of deformable anatomic structure," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[98] D. Vilkomerson, S. Ricci, and P. Tortoli, "Finding the peak velocity in a flow from its doppler spectrum," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 60, no. 10, pp. 2079–2088, 2013.

[99] S. Ricci, R. Matera, and P. Tortoli, "An improved doppler model for obtaining accurate maximum blood velocities," *Ultrasonics*, vol. 54, no. 7, pp. 2006–2014, 2014.

[100] A. Kathpalia, Y. Karabiyik, S. H. Eik-Nes, E. Tegnander, I. K. Ekroll, G. Kiss, and H. Torp, "Adaptive spectral envelope estimation for doppler ultrasound," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 11, pp. 1825–1838, 2016.

[101] Q. He, H. Peng, Z. Han, C. Zheng, and Y. Wang, "A new algorithm for blood flow measurement based on the doppler flow spectrogram," *IEEE Access*, vol. 7, pp. 468–477, 2019.

[102] C. Fancourt, K. Azer, S. L. Ramcharan, M. Bunzel, B. R. Campbell, J. R. Sachs, and M. Walker, "Segmentation of arterial vessel wall motion to sub-pixel resolution using m-mode ultrasound," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3138–3141, IEEE, 2008.

[103] K. E. Leung and J. G. Bosch, "Localized shape variations for classifying wall motion in echocardiograms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 52–59, Springer, 2007.

[104] M. Qazi, G. Fung, S. Krishnan, R. Rosales, H. Steck, R. B. Rao, D. Poldermans, and D. Chandrasekaran, "Automated heart wall motion abnormality detection from ultrasound images using bayesian networks.," in *IJCAI*, vol. 7, pp. 519–525, 2007.

[105] A. Shalbaf, H. Behnam, Z. Alizade-Sani, and M. Shojaifard, "Automatic classification of left ventricular regional wall motion abnormalities in echocardiography images using nonrigid image registration," *Journal of digital imaging*, vol. 26, no. 5, pp. 909–919, 2013.

[106] T. Araki, N. Ikeda, D. Shukla, N. D. Londhe, V. K. Shrivastava, S. K. Banchhor, L. Saba, A. Nicolaides, S. Shafique, J. R. Laird, *et al.*, "A new method for ivus-based coronary artery disease risk stratification: a link between coronary & carotid ultrasound plaque burdens," *Computer methods and programs in biomedicine*, vol. 124, pp. 161–179, 2016.

[107] S. G. Mougiakakou, S. Golemati, I. Gousias, A. N. Nicolaides, and K. S. Nikita, "Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image statistics, laws' texture and neural networks," *Ultrasound in medicine & biology*, vol. 33, no. 1, pp. 26–36, 2007.

[108] U. R. Acharya, M. R. K. Mookiah, S. V. Sree, D. Afonso, J. Sanches, S. Shafique, A. Nicolaides, L. M. Pedro, J. F. e Fernandes, and J. S. Suri, "Atherosclerotic plaque tissue characterization in 2d ultrasound longitudinal carotid scans for automated classification: a paradigm for stroke risk assessment," *Medical & biological engineering & computing*, vol. 51, no. 5, pp. 513–523, 2013.

[109] U. Raghavendra, H. Fujita, A. Gudigar, R. Shetty, K. Nayak, U. Pai, J. Samanth, and U. R. Acharya, "Automated technique for coronary artery disease characterization and classification using dd-dtdwt in ultrasound images," *Biomedical Signal Processing and Control*, vol. 40, pp. 324–334, 2018.

[110] B. Smitha and K. P. Joseph, "A new approach for classification of atherosclerosis of common carotid artery from ultrasound images," *Journal of Mechanics in Medicine and Biology*, vol. 19, no. 01, p. 1940001, 2019.

[111] V. K. Sudarshan, U. R. Acharya, E. Ng, R. San Tan, S. M. Chou, and D. N. Ghista, "Data mining framework for identification of myocardial infarction stages in ultrasound: A hybrid feature extraction paradigm (part 2)," *Computers in biology and medicine*, vol. 71, pp. 241–251, 2016.

[112] K. S. Vidya, E. Ng, U. R. Acharya, S. M. Chou, R. San Tan, and D. N. Ghista, "Computer-aided diagnosis of myocardial infarction using ultrasound images with dwt, glcm and hos methods: a comparative study," *Computers in biology and medicine*, vol. 62, pp. 86–93, 2015.

[113] V. K. Sudarshan, U. R. Acharya, E. Ng, R. San Tan, S. M. Chou, and D. N. Ghista, "An integrated index for automated detection of infarcted myocardium from cross-sectional echocardiograms using texton-based features (part 1)," *Computers in biology and medicine*, vol. 71, pp. 231–240, 2016.

[114] G. Balaji, T. Subashini, and N. Chidambaram, "Detection and diagnosis of dilated cardiomyopathy and hypertrophic cardiomyopathy using image processing techniques," *Engineering Science and Technology, an International Journal*, vol. 19, no. 4, pp. 1871–1880, 2016.

[115] S. Narula, K. Shameer, A. M. S. Omar, J. T. Dudley, and P. P. Sengupta, "Machine-learning algorithms to automate morphological and functional assessments in 2d echocardiography," *Journal of the American College of Cardiology*, vol. 68, no. 21, pp. 2287–2295, 2016.

[116] K. K. Shung, *Diagnostic ultrasound: Imaging and blood flow measurements*. CRC press, 2015.

[117] O. Faust, U. R. Acharya, V. K. Sudarshan, R. San Tan, C. H. Yeong, F. Molinari, and K. H. Ng, "Computer aided diagnosis of coronary artery disease, myocardial infarction and carotid atherosclerosis using ultrasound images: A review," *Physica Medica*, vol. 33, pp. 1–15, 2017.

[118] M. Negahdar, M. Moradi, N. Parajuli, and T. Syeda-Mahmood, "Automatic extraction of disease-specific features from doppler images," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 101340N, International Society for Optics and Photonics, 2017.

[119] P. C. Tang, A. Nicoara, and C. A. Milano, "Tricuspid valve regurgitation and right ventricular dysfunction during left ventricular assist device implantation," in *Mechanical Circulatory Support in End-Stage Heart Failure*, pp. 221–226, Springer, 2017.

[120] D. E. Bild, D. A. Bluemke, G. L. Burke, R. Detrano, A. V. Diez Roux, A. R. Folsom, P. Greenland, D. R. JacobsJr, R. Kronmal, K. Liu, *et al.*, "Multi-ethnic study of atherosclerosis: objectives and design," *American journal of epidemiology*, vol. 156, no. 9, pp. 871–881, 2002.

[121] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.

[122] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, 2009.

[123] A. H. Kadish, D. Bello, J. P. Finn, R. O. Bonow, A. Schaechter, H. Subacius, C. Albert, J. P. Daubert, C. G. Fonseca, and J. J. Goldberger, "Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (determine) trial," *Journal of cardiovascular electrophysiology*, vol. 20, no. 9, pp. 982–987, 2009.

[124] A. Suinesiaputra, D. A. Bluemke, B. R. Cowan, M. G. Friedrich, C. M. Kramer, R. Kwong, S. Plein, J. Schulz-Menger, J. J. Westenberg, A. A. Young, *et al.*, "Quantification of lv function and mass by cardiovascular magnetic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

22

resonance: multi-center variability and consensus contours," *Journal of Cardiovascular Magnetic Resonance*, vol. 17, no. 1, p. 63, 2015.

[125] K. Gilbert, N. Forsch, S. Hegde, C. Mauger, J. H. Omens, J. C. Perry, B. Pontré, A. Suinesiaputra, A. A. Young, and A. D. McCulloch, "Atlas-based computational analysis of heart shape and function in congenital heart disease," *Journal of cardiovascular translational research*, vol. 11, no. 2, pp. 123–132, 2018.

[126] C. Tobon-Gomez, A. J. Geers, J. Peters, J. Weese, K. Pinto, R. Karim, M. Ammar, A. Daoudi, J. Margeta, Z. Sandoval, *et al.*, "Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets," *IEEE transactions on medical imaging*, vol. 34, no. 7, pp. 1460–1473, 2015.

[127] D. Ouyang, B. He, A. Ghorbani, C. Langlotz, P. A. Heidenreich, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Interpretable ai for beat-to-beat cardiac function assessment," *medRxiv*, p. 19012419, 2019.

[128] O. Bernard, J. G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea, M. Bernier, *et al.*, "Standardized evaluation system for left ventricular segmentation algorithms in 3d echocardiography," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 967–977, 2016.

[129] C. Tobon-Gomez, M. De Craene, K. Mcleod, L. Tautz, W. Shi, A. Hennemuth, A. Prakosa, H. Wang, G. Carr-White, S. Kapetanakis, *et al.*, "Benchmarking framework for myocardial tracking and deformation algorithms: An open access database," *Medical image analysis*, vol. 17, no. 6, pp. 632–648, 2013.

[130] R. Robinson, V. V. Valindria, W. Bai, O. Oktay, B. Kainz, H. Suzuki, M. M. Sanghvi, N. Aung, J. M. Paiva, F. Zemrak, *et al.*, "Automated quality control in image segmentation: application to the uk biobank cardiovascular magnetic resonance imaging study," *Journal of Cardiovascular Magnetic Resonance*, vol. 21, no. 1, p. 18, 2019.

[131] J. Rychik, "Fetal cardiovascular physiology," *Pediatric cardiology*, vol. 25, no. 3, pp. 201–209, 2004.

[132] Q. Zhang, Y. Nian Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8827–8836, 2018.

[133] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[134] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2d echocardiography," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2198–2210, 2019.

[135] S. Dong, G. Luo, K. Wang, S. Cao, Q. Li, and H. Zhang, "A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3d echocardiography," *BioMed research international*, vol. 2018, 2018.

[136] D. Krishnaswamy, A. R. Hareendranathan, T. Suwatanaviroj, H. Becher, M. Noga, and K. Punithakumar, "A semi-automated method for measurement of left ventricular volumes in 3d echocardiography," *IEEE Access*, vol. 6, pp. 16336–16344, 2018.

[137] M. Bernier, P.-M. Jodoin, O. Humbert, and A. Lalande, "Graph cut-based method for segmenting the left ventricle from mri or echocardiographic images," *Computerized Medical Imaging and Graphics*, vol. 58, pp. 1–12, 2017.

[138] X. Huang, D. P. Dione, C. B. Compas, X. Papademetris, B. A. Lin, A. Bregasi, A. J. Sinusas, L. H. Staib, and J. S. Duncan, "Contour tracking in echocardiographic sequences via sparse representation and dictionary learning," *Medical image analysis*, vol. 18, no. 2, pp. 253–271, 2014.

[139] M. De Craene, S. Marchesseau, B. Heyde, H. Gao, M. Alessandrini, O. Bernard, G. Piella, A. Porras, L. Tautz, A. Hennemuth, *et al.*, "3d strain assessment in ultrasound (straus): A synthetic comparison of five tracking methodologies," *IEEE transactions on medical imaging*, vol. 32, no. 9, pp. 1632–1646, 2013.

APPENDIX A
ECHOCARDIOGRAPHY MODES

A.1 M-mode

M-mode provides a one dimensional view or trace of the motion for a specific cardiac structure (e.g., mitral valve). It helps visualizing the temporal changes in the depth of echo-producing interfaces. The X axis (top or bottom) of M-mode shows time and the Y axis (sides) shows distance. This type of echo imaging has a high temporal resolution and is useful for measuring rapid motions (e.g., opening and closing of valves). Figure 2 presents an example of an image acquired using M-mode echo.
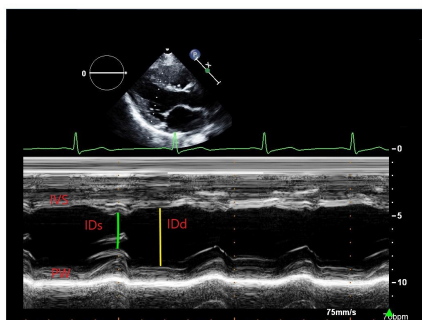

Fig. 3. B-mode A4C view echocardiography.

Fig. 2. M-mode echocardiography recorded from the first author. IVS: Inter-ventricular Septum; PW: Posterior Wall; IDs: Internal Diameter (systole); IDd: Internal Diameter (diastole); green curve is ECG signal.

Fig. 4. Spectral Doppler mode echocardiography (mitral valve [MV] flow). The upper part of the image shows the color Doppler. Image is taken from Open-i.

A.2 B-mode

This mode provides a cross-sectional image of the heart's tissues and boundaries. Each point in B-mode represents an echo and the brightness (B-mode) of each point represents the strength of the reflected echo. A comprehensive B-mode acquisition involves imaging the heart from different windows or views by positioning the transducer in different locations [2]. The most common B-mode views include [2]: parasternal long axis and short axis (PLAX and PSAX), apical two-chamber (A2C), apical three-chamber (A3C), apical four-chamber (A4C), apical five-chamber (A5C), Sub-Costal Long and Short Axis View (SCLX and SCSX), and Suprasternal Notch View (SSN). Figure 3 presents an example of B-mode A4C view.

A.3 Doppler mode

Doppler measures the velocity and direction of blood cells within the heart. There are two main types of Doppler imaging: *Color Doppler* and *Spectral Doppler*. *Color Doppler* visualizes blood flow direction and velocity using a color scale, where red hues represent flow toward and blue hues represent flow away from the transducer. As shown in Figure 4, *color Doppler* is usually superimposed on B-mode grey-scale image.
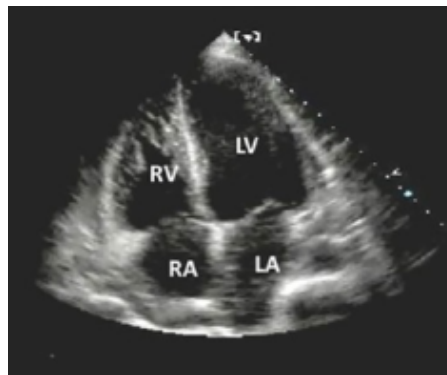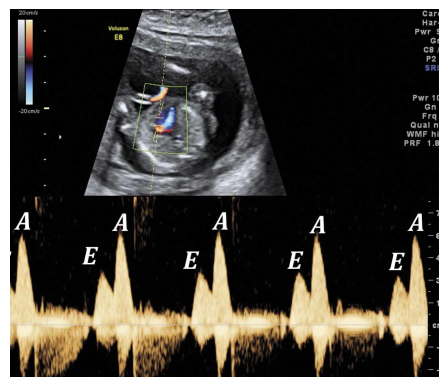
Spectral Doppler (Figure 4) uses the frequency shift in reflected waves to visualize the blood flow as a graph that shows the velocity of blood flow (Y axis) over time (X axis). A velocity value displayed above the baseline indicates flow towards the transducer and a value displayed below the baseline indicates flow away. The baseline is a horizontal line that has zero velocity. This type of echocardiography is routinely performed using either continuous or pulsed wave Doppler [2]. *Pulsed Wave Doppler* or PW Doppler utilizes a single transducer element to send and receive an ultrasound wave. By sending and receiving pulses, PW Doppler has the ability to measure the velocity of blood flow at a specific cardiac region (a.k.a., sample volume). This makes PW Doppler a very powerful method for providing site-specific information. However, a major limitation of PW Doppler is its inability to display high velocities due to aliasing phenomena [2]. *Continuous Wave Doppler* or CW Doppler, on the other hand, accurately measures high blood velocities. It has two dedicated transducer elements for continuously sending and receiving ultrasound waves. This type of Doppler is not site-specific and is frequently used to measure high blood velocities of cardiac pathologies [2].

TABLE 2
Quantitative comparison of automated methods for echo quality assessment. A2C (apical 2 view), A3C (apical 3 view), A4C (apical 4 view), $PLAX_A$ (parasternal long axis view, aortic valve), $PLAX_{PM}$ (parasternal long axis view, papillary muscle), GHT (Generalized Hough Transform), CNN (Convolutional neural network), TPR (true positive rate), CC (correlation coefficient).

Work	ROI Method	Mode & View	Method	System & Data	Train & Test	Ground Truth	Performance
[27]	NA	B-mode: A4C	**Model-based:** B-splines to model four chambers; goodness-of-fit	GE Vivid E9 system 95 videos	Train: 4 patients 35 cases Test: 2 patients 60 cases	Scores by 2 cardiologists: Good, fair, and poor	TPR (Section 3.1): Good quality: 22% Fair quality: 20% Poor quality: 15%
[28]	NA	B-mode: PLAX	**Model-based:** GHT applied to input image compared with Atlas: created from images segmented manually	GE Vivid 7 system 133 images 35 patients	Train: 89 images to create PLAX Atlas Test: 44	Scores by expert sonographer: Good (score 3) Poor (score 0)	CC (Section 3): 0.84 correlation between manual and automated scores
[29]	NA	B-mode: A4C	**Deep Learning:** Customized regression CNN	NA system; 2,904 A4C images	Train: 80% 2,345 images; Test: 20% 560 images;	Scores by expert cardiologist: Good and Poor	Mean Absolute Error (MAE): 0.87 ± 0.72
[30]	NA	B-mode: AP2, AP3, AP4, $PLAX_A$, $PLAX_{PM}$	**Deep Learning:** Customized regression CNN	Different GE and Philips systems; 2,450 cines: A2C (478), A3C (455), A4C (575), $PLAX_A(480)$, $PLAX_P(462)$	Train: 80% # videos per view = 935 Total (4,675); Test: 20% # videos per view = 228 Total (1,144); 20 frames videos	Scores by physicians: A2C (0-8), A3C (0-7), A4C (0-10), $PLAX_A(0-4)$, $PLAX_{PM}$ (0-5), scores normalized	View accuracy: $(1 - \Sigma_1^T \lvert A - M \rvert)$ T: cases per view A-M: auto-hand, A2C (86 ± 9), A3C (89 ± 9), A4C (83 ± 14), $PLAX_A(84 \pm 12)$, $PLAX_P(83 \pm 13)$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

25

TABLE 3
Quantitative comparison of automated methods for mode/view classification. A2C-A3C-A4C-A5C (apical 2-3-4-5 views), PLAX (parasternal long axis view), PSAX (parasternal short axis view), PSA (Parasternal Short Axis), PSAM-PASAP (PSA of Mitral and Papillary), IVC (inferior vena cava), SC2C-SC4C-SCLX (subcostal 2-4 chamber and long axis), GSAT (Gray- Level Symmetric Axis Transform), SVM (support vector machines), TPR (True Positive Rate).

Work	ROI Method	Mode & View	Method	System & Data	Train & Test	Ground Truth	Performance
[32]	NA	B-mode: A2C, A4C, PALX, PSAX, SC2C, SC4C, SCLX, other	**Conv. ML method** GIST descriptor, probabilistic SVM	Philips CX50; 33Hz; 270 videos, 5-10 heartbeats	Train: 2700, Test: 2700 frames	Domain expert annotation for all views	TPR (Section 3.1): A2C-A4C (100), SAX (100), LAX (98), SC2 (96), SC4 (100), SCL (64), other (96)
[35]	LV Detectors (MLBoost) in all viwes	B-mode: A2C, A4C, PALX, PSAX	**Conv. ML method** Fusion of LV detectors; multi-class boosting	System: NA; 1303 videos, A2C (371), A4C (574), PSAX (203), PLAX (155)	Train: 1080; Test: A2C (61), A4C (96), PSAX (28), PLAX (38)	Manually localized LV regions	TPR (Section 3.1): A2C (93.5), A4C (97.9), PSAX (96.4), PLAX (97.4)
[36]	GSAT Detector	B-mode: A4C, PALX, PSAX	**Conv. ML method** Relational Structures, Markov Random Field, multi-class SVM	System: NA; 15 normal vid., 2657 i-frames; 6 abnormal, 552 i-frames;	Train: 2657, leave-one-out; Test: 552	Domain expert annotation for all views	Average precision (Section 3.1): 88.35%
[37]	Manual	B-mode: A2C, A3C, A4C, A5C, SAB, SAP, PLA, PSAM	**Conv. ML method** Optical flow, edge-filtered map, SIFT features, SVM	System: NA 113 vid., 25 Hz 320 × 240 pix. 2470 frames	leave-one-out;	Manual labeling	TPR (Section 3.1): A2C (51), A3C (54), A4C (93), A5C (61), SAB (1.0), SAP (93), PLA (88), PSAM (71)
[46]	NA	B-mode: A2C, A3C, A4C, PLAX, PSAX, IVC, other	**Deep learning:** VGG-based CNN with 6 classes; ADAM, 64 batch $1x10^5$ learning rate, 10-20 epochs; 2 hr training (GTX 1080), 600 ms runtime	System: NA > 4000 studies	Train: 40,000 images; Test: VC (159), A2C (555), A3C (174), A4C (756), PLAX (515), PSAX (458)	Manual labeling	TPR (Section 3.1): IVC (100), A2C (94) A3C (93), A4C (98), PLAX (99), PSAX (99.5)
[49]	NA	B-mode: A2C, A4C, PLAX, PSAX, ALAX, SC4C, SCVC, unknown	**Deep learning:** Inception-based CNN with 7 classes; Adam, 10^{-4} rate, 64 mini-batch, 100 epochs	GE Vivid E9, 4582 vid., 205 patients, avg. age: 64; GE Vivid E7, 2559 vid., 265 patients, avg. age: 49	Train: 4582 vids., 256,649 frames; Test: 2559 vids., 229,951 frames	Manual labeling	Overall accuracy: Frame (98.3 ± 0.6) Video (98.9 ± 0.6); runtime (4.4 ± 0.3) ms (GPU)
[50]	NA	B-mode: 12 apical, parasternal, subcostal, suprasternal views	**Deep learning:** Lightweights VGG, DenseNet, and ResNet based models; ADAM, 1^{-4} rate, 300 batch	Philips, GE, and Siemens systems; 3,151 patients, 16,612 cines, 807908 frames	Patient level split: Train (60%), Valid (20%), Test (20%),	Manual labeling by senior cardiologist	Overall accuracy: 88.1%; fusion of 3 models
[51]	NA	B-mode: A2C, A3C, A4C, A5C, PLAX, PSAX, PSAM, PSAP	**Deep learning:** Spatial CNN, input: raw image; Temporal CNN, input: acceleration image	GE Vivid 7 or E9; 432 vid.; age: 7-85; 434 × 636, 26fps 341 × 415, 26fps	Train: 280, Test: 152; Re-sized: 227 × 227× 26 frames	Clinicians in 2 hospitals labeled 8 views	TPR (Section 3.1): A2C:100, A3C:100, A4C:100, A5C:71.4, PLAX:96, PSAX:95, PSAM:88, PSAP:75

TABLE 4

Quantitative comparison of state-of-the-art segmentation methods. LV (Left Ventricle), A2C-A4C (apical 2-4 view), PLAX (parasternal long axis view), PSAX (parasternal short axis view), CW (continuous wave), PW (pulse wave), MV (mitral valve), AV (aortic valve), ED (end-diastole), ES (end-systole), RMSE (root mean square error), SRAD (speckle reduction anisotropic diffusion), PPG (peak pressure gradient), PHT (pressure half time), EPV (E-Wave Peak Velocity), EDT (E-Wave deceleration time), APV (A-Wave Peak Velocity), and ADU (A-Wave duration).

** indicates the performance is reported by superimposing the automated segmentation on raw images. We refer the reader to the actual papers for visualization of the results as including these images would require obtaining permission from the publisher.

Work	ROI Method	Mode & View	Method	System & Data	Ground Truth	Performance
[52]	Semi-automated	B-mode: A4C, LV	**Low Level Image Processing** Pre-processing module (Filtering & morphological operations) Segmentation module (Watershed and contour correction)	ATL HDI 3000; 12 volunteers, 12 vid. 44fps, 900 frames	Manual contours by a specialist	CC (Section 3.2): 0.87 High (0.99±0.01) Average (0.90±0.024) Low (0.73±0.101)
[54]	Automated (k-means)	B-mode: A4C, all chambers	**Low Level Image Processing** SRAD filtering, thresholding, edge detection	System: NA; 20 volunteers, 25 videos	Manual contours by a specialist	**
[60]	NA	B-mode: A2C, LV	**Deformable Model:** Active contour, coupled optimization function	System: NA; 61 volunteers, 85 ED images	Manual contours for 85 images by expert echo	**
[61]	NA	B-mode: PLAX, PSAX LV	**Deformable model:** Smoothing, Hough transform, active contour	System: NA; 11 volunteers, 15 ED images	Manual contours for 85 images by echo expert	**
[62]	Manual	B-mode: A2C, A4C LV	**Deformable Model:** Control points located manually (initial contour) B-spline snake (final contour)	System: GE ; Vivid 3; 50 ES and ED images	Manual contours and cardiac indices by echo expert	RMSE between auto and manual: LV area: 1.5, LV volume: 6.8, Ejection fraction: 4.6
[72]	Manual	B-mode: A4C, LV LV	**Statistical Model:** Global despeckling, active appearance model training,	System: NA; synthetic and clinical echo images, 56 normal fetuses	Manual contours by cardiologist	Pixel accuracy (Section 3.1): Synthetic (84.12%), Clinical (84.39%)
[76]	NA	B-mode: A4C, all chambers	**Conventional Machine Learning:** Adaptive Group Dictionary Learning, Dictionary initialization, sparse group representation, pixel classification	System: NA; 40 clinical images of 50 normal fetuses	Manual contours by cardiologist	**
[46]	NA	B-mode: A2C, A4C PLAX, PSAX; all chambers	**Deep Learning (Pixel Segmentation):** 4 U-net CNN models trained using images and masks (A2C = 198, A4C = 168, PSAX = 72, PLAX = 128); Augmentation (cropping & blackout); Training, 2 hours on Nvidia GTX 1080; Runtime: 110ms per image on average	System: NA; Train: 566 images and masks; Test: 557 images	Manual segmentation of all chambers	** IOU value (Section 3.2): 55% to 92% for all views and chambers
[82]	Manual	Spectral Doppler; long strips	**Low Level Image Processing:** Objective thresholding method, morphological operations, biggest-gap algorithm for peak detection	GE Vivid 5; 25 CW & PW normal images	Manual velocity time integral & peak-velocity by cardiologist	CC (Section 3.2): velocity-integral (0.94) peak-velocity (0.98)
[83]	Detection based on axes fixed locations	Spectral CW Doppler;	**Low Level Image Processing:** Noise filtering & contrast adjustment, Canny edge detector, envelope smoothing, peak detector smoothing, peak detector	System: NA; 22 images; 11 normal subjects; 3 age groups	Manual peak velocity, PPG, and PHF by a cardiologist	CC (Section 3.2): Age G1 (20-35): 0.985, Age G2 (36-50): 0.922, Age G3 (51-65): 0.833
[84]	Manual	Spectral CW Doppler	**Low Level Image Processing:** Texture filters (entropy, range, and standard deviation), thresholding, morphological operations	System: NA; 20 CW images; 25 patients with AR	Manual envelope contours by a cardiologist	**
[63]	NA	Spectral Doppler, MV & AV	**Deformable Model:** Speckle resistant gradient vector flow, Generalized gradient vector flow field	Philips devices; 30 patients, 10 with atrial fibrillation, 20 normal	Manual velocity time integral, peak velocity, & border contours by 2 experts	CC and B&A (Section 3.2): see [63] for complete results
[93]	NA	Spectral CW Doppler	**Model-based:** Reference image calculated from all training images (model), mapping or registration from input to reference	GE Vivid 7; 59 CW images; 30 normal volunteers	Manual envelope delineation by echo expert	**
[96]	Automated; 3 trained detectors	Spectral Doppler: MV	**Conventional Machine Learning:** E peak detector (left root), A velocity detector (right root), peak detector; training shape inference model (mapping from image to its shape)	System: NA; 255 training, 43 testing	Manual Doppler indices (EPV, EDT, (APV, ADU) by 2 sonographers	CC (Section 3.2): EPV (0.987), EDT (0.821), APV (0.986), EDU (0.481)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2988295, IEEE Reviews in Biomedical Engineering

27

TABLE 5
Summary of automated CVD classification methods for B-mode and Doppler. LV (left ventricle), WMA (wall motion abnormalities), A2C-A3C-A4C (apical 2-3-4 view), CW (continuous wave), CAD (coronary artery disease), DCM (dilated cardiomyopathy), HCM (Hypertrophic cardiomyopathy), ATH (physiological hypertrophy in athletes), MI (myocardial infarction), AS (aortic stenosis), AR (aortic regurgitation), GLCM (gray level co-occurrence matrix), GLRLM (gray level run length matrix), GLDS (gray level difference statistics), SM (statistical feature matrix), LCP (Local Configuration Pattern), PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), DCT (Discrete cosine transform), DT (Decision Tree), RF (Random Forest), NN (neural network), SVM (support vector machine), k-NN (k-nearest neighbors), IG (information gain), KS-test (Kolmogorov-Smirnov test), mRMR (Max-Relevancy and Min-Redundancy Feature).

Work	ROI	Objective	Data & Labels	Method	Features & Markers	Top Features	Performance
[103]	NA	LV WMA Detection; A2C, A4C B-mode	129 patients, 65 patients (train), 64 patients (test); LV contours and Abnormalities scores by 2 expert readers	LV modeling using PCA; shape modes describe variations in the population; LDA classifier	Features: statistical parameters extracted from shape models; Biomarkers: NA	8 PCA parameters;	Avg. accuracy (Correctly classified cases): 88.9
[104]	NA	LV WMA Detection; A2C, A3C, A4C B-mode	Data of normal & abnormal (hypokinetic, akinetic, dyskinetic, ,aneurysm) patients; 220/125, train/test Abnormalities scores	Hand-initialized dual-contours (endocardium and epicardium) tracked over time; bayesian networks (binary)	Features extracted from contour: circumferential strain, radial strain, local, global, and segmental volume markers	6 features (global & local) based on KS-test	Sensitivity (Section 3.1): 80 to 90
[105]	Manual	LV WMA Detection; A2C & A4C B-mode	Data of 10 healthy & and 14 patients with ischemic; 336 segments: 55% normal, 13% hypo -kinetic, 31% akinetic; 220/125, train/test; Abnormalities scores	Affine registration and B-spline snake to model LV; threshold classifier	Novel regional index computed from control points of B-spline snake; Biomarkers: NA	New Quantitative Regional Index	Agreement between 2 experts and automated: Absolute, 83 Relative, 99
[106]	Manual	CAD risk assessment; B-mode	Stroke-risk (>0.9mm) to label patients as: High risk CAD (9), Low risk CAD (6); 1508 frames high risk, 1357 frames low risk; ROIs by 2 experts	56 grayscale feature extracted: GLCM, GLRLM, GLDS, SM, invariant moment; SVM classifier; k-fold cross validation	Derived 6 Feature Combinations: FC1, FC2, FC3, F4, F5, F6; Biomarkers: NA	Best feature set was chosen based on classification accuracy (FC6)	Avg. accuracy (Section 3.1): 94.95; AUC: 0.95;
[111]	NA	MI stage detection; A4C B-mode	WMSI & LVEF to label patients as: normal (40), 200 moderate (40), severe (40); 600 images, 200 per class; age: 21-75	Curvelet Transform and LCP features; LDA, SVM, DT, NB, kNN, NN for classification; 10-fold cross validation	17,850 LCP features extracted from 46,200 CT coefficients; Biomarkers: NA	mRMR method: 30 coefficients, 6 features; proposed Myocardial Infarction Risk Index (MIRI)	Accuracy: 98.99; sensitivity: 98.48; specificity: 100% (SVM, RBF)
[114]	Auto. Fuzzy c-means (FCM)	DCM & HCM detection; LV, PSAX, B-mode	Data of 20 normal, 30 DCM, and 10 HCM patients; 60 (4-6 seconds) videos, 46 fps	LV segmentation by FCM clustering; shape & statistical (PCA & DCT) features; NN, SVM & combine k-NN for classification	DCT & PCA features; Biomarkers: EF, EDV, ESV, mass, septal thickness	PCA features is better than DCT and LV biomarkers	TPR: 92.04 (normal, abnormal) (NN)
[115]	NA	Distinguish HCM & ATH; LV, A4C, B-mode	139 male subjects, 77 with ATH, 62 with HCM; poor quality images excluded	TomTec software for LV speckle tracking; ensemble of NN, SVM, RF for classification; 10 cross validation	Speckle-tracking based geometric (e.g., volume) & mechanical (e.g., velocity) parameters	Based on info. gain (IG): Volume (0.24), MLVS (0.134), ALS (0.13)	Sensitivity: overall (87), adjusted for age (96); Specificity: overall (82), adjusted for age (77)
[83]	NA	AR assessment; CW, Doppler	9 male & 2 female subjects with mild, moderate, severe AR; 22 images; 3 age groups: G1 (20-35), G2 (36-50), G3 (51-65); ground truth by experts	Envelope delineation: filtering, morpho-logical operations, thresholding, edge detection	Parameters computed from detected envelope: peak velocity, pressure gradient, pressure half time	Pressure half time (PHT)	High CC between automated and manual: r=0.95
[63]	NA	Valves dysfunctions quantification; CW, Doppler	60 patients: 30 with aortic/mitral stenosis; 20 with normal sinus rhythm; 10 with atrial fibrillation; ground truth: manual indices by expert	Envelope delineation: Active contour for envelope delineation	Doppler indices computed from detected envelope: Peak velocity (P_V), Mean velocity (M_V), Velocity time integral (VTI)	Mean velocity (M_V)	B & A, LOA (Section 3.2): (-3.9 to +0.5), (-4.6 to -1.4), (-3.6 to +4.4) for P_V, M_V, and VTI (acceptable)

TABLE 6
The characteristics of publicly available echocardiography datasets. HF (heart failure), CAD (coronary artery disease), DCM (dilated cardiomyopathy), MI (myocardial infarction), SD (standard deviation), LV (left ventricle), EF (ejection fraction), ESV (end-systole volume), EDV (end-diastole volume).

Dataset	Subjects	Location	System & Data	Exclusion	Train/Test	Ground Truth	Recent Works
EchoNet Dynamic [127]	10,025 patients; 49% female; average age 68 ± 21; 29% HF patients, 23% CAD patients	Stanford University Hospital (2006 to 2018)	Acuson SC2000, Siemens, Epiq 5G, Epiq 7, Philips; 10,025 2D A4C videos, 30 FPS	NA	Train: 7465 Valid.: 1288, Test: 1277 patients	EF, ESV, EDV indices; tracing at ES and ED frames (LV)	[127]
CETUS [128]	45 subjects: 15 healthy, 15 with MI, 15 with DCM	University Hospital France; University Hospital Leuven-Belgium, Erasmus; University MC-Netherlands	GE, Vivid E9, 4V probe; Philips, iE33, X5-1 probe; Siemens, SC2000, 4Z1c probe; 3D (A4C) videos, 25.7 ± 8.5 mean ± SD frames per cycle	Visually dys-synchronous LV; unacceptable quality image (labeled by cardiologists)	Train: 15, Test: 30 patients	Manual LV contours provided by 3 cardiologists; marked ED and ES frames	[134] [135] [136] [137]
STACOM [129]	16 anatomies: 15 volunteers, 1 phantom; 3 female, aged 28 ± 5 year; 2 modalities: Ultrasound and MRI	Healthy: Division of Imaging Sciences and Biomedical Engineering, King's College London; Phantom: Department of Internal Medicine Cardiology, University of Ulm	13-30 time frames 1158 image volumes	Unacceptable quality images	NA	12 points tracked by 2 observers and registered to 3D coordinates using a point based similarity transform; Quality scores provided by 2 cardiologists	[138] [139]